This electronic thesis or dissertation has been downloaded from the King's Research Portal at https://kclpure.kcl.ac.uk/portal/

Processing of Cardiac Signals for Health Monitoring and Early Detection of Heart Diseases

Bao, Xinqi

Awarding institution:
King's College London

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

END USER LICENCE AGREEMENT

Unless another licence is stated on the immediately following page this work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International licence. https://creativecommons.org/licenses/by-nc-nd/4.0/

You are free to copy, distribute and transmit the work

Under the following conditions:

- Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
- Non Commercial: You may not use this work for commercial purposes.
- No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 25. Dec. 2024

Processing of Cardiac Signals for Health Monitoring and Early Detection of Heart Diseases

Xinqi Bao

Supervisor: Dr. Ernest Kamavuako
Dr. Yansha Deng

The Department of Engineering
King's College London

This dissertation is submitted for the degree of
Doctor of Philosophy

April 2023

Acknowledgements

Life is like a journey, and my PhD is also a significant adventure. During this period, I had the privilege to explore and make my own contribution in the field of science and technology. Besides, the experience of living abroad independently made me grow up gradually. Although it feels like yesterday when I first arrived and started my project, farewell always exists. Now, it is the time to summarize and prepare to embark on a new journey. I dedicate this essay to all the people who helped and stood behind me. Without you, it cannot be a success.

Although the 4-year PhD life was colorful, most of the academic work was still stressful and challenging. So first of all, I would like to thank my supervisor Dr. Ernest. He guided me patiently with wisdom to find the rhythm of research and helped me develop academic skills. In addition, he gave me enough trust and opportunities to manage the laboratory, participate in teaching, attend academic conference, try the entrepreneurship, and cultivate comprehensive capabilities. These will benefit me for life.

The life of these years was not always smooth. The successful completion of my PhD was also credited to the understanding and help of my family and friends. Therefore, I firstly appreciate my parents that they can always support me firmly and be my solid backing no matter what decision I made. I then thank my flatmates who are also my colleagues, Dr. Fenghe and Yujia, to provide understanding and tolerance in my daily life, and also inspire me in study. Dr. Xianqi, Yuzhi, Mengqi, Shilei and other PhD friends, thank you all for always inviting me in holidays, so that I did not feel alone. Also thank my lovely cat Matcha, who provided me a lot of happiness during the pandemic and always released my anxiety when I worked late at night. These memories I will never forget. Additionally, all the academics, colleagues and friends who worked and exchanged with me were important components of my PhD life. Due to space limitations, cannot name one by one. But I sincerely hope the friendship will last forever.

Where there are gains, there are also losses. For the people who were important and have gradually disappeared in my life, thank Dr. Zhaozhang, you motivated me to do PhD and made who I am now; thank Dr. Anqi, when facing the bottlenecks of research and fatigue of life during the pandemic, you gave me mental support and encouragement to persist in making myself better; thank Mingyi, your companionship has made me feel the existence and enthusiasm in the simple life by rules.

Though I have experienced a lot of confusion and trough over the years, the memories are all about sense of accomplishment and pride. The life journey still continues, and all the experience and growth of these years will give me the capability and courage to face all the pressures and challenges in future work and life.

Abstract

Cardiovascular disease (CVD) is the leading cause of mortality, accounting for 30% of deaths worldwide. Early screening and real-time monitoring play a vital role in the detection and taking necessary action to reduce the risk of worsening heart disease. The initial suspicion often depends on the medical staff to listen to murmurs in the heart sound (recorded as phonocardiogram, PCG) during auscultation, a sign of deflection from the electrocardiogram (ECG) and other signs such as low oxygen saturation (SpO2) and changes in respiration. However, these screening methods rely heavily on the physician's auscultation experience, the ability to interpret ECG signals, and the real-time monitoring of multiple physiological signals. In the last decade, the rapid development of wearable devices and machine learning, especially deep learning techniques, has enabled the miniaturised and portable CVD screening devices in primary care.

Nonetheless, the existing devices mainly provide single physiological signal measurement and cannot perform reliable assistive diagnoses, which limits the applicability of these devices. As a result, it is of great value to integrate multiple signal measurements with diagnostic capabilities on miniaturised screening devices in the future. For long-term vision, the research aims at designing a multi-sensor miniaturised device for cardiac investigation and monitoring, however, this thesis aims at (1) developing machine learning techniques to improve the computer-aided diagnosis of PCG and ECG; (2) assessing the feasibility by experiments, utilising biomedical signal processing to eliminate the need of physical respiration sensor. Specifically, the proposed objectives and outcomes are as follows:

1) To assess the feasibility of localised ECG signal acquisition and analyse its usability for PCG segmentation. We experimentally investigated the time property of ECG and PCG signals at auscultation sites and the effect of ECG inter-electrode distance. Results showed that ECG signal could be acquired stably at auscultation sites within a small area (5 cm), which provides a theoretical basis for designing miniaturised integrated ECG-PCG devices. Furthermore, the accuracy and robustness of PCG segmentation have always been important issues affecting PCG recognition. The obtained temporal relationships in this study will also make the device to perform reliable PCG segmentation using ECG signals.

2) To investigate the optimal use of deep learning input and propose a reliable algorithm for PCG classification. The conducted research aimed at optimising the information to improve the classification performance of deep learning. On recurrent neural networks (RNNs), the study analysed input length's influence on classification accuracy using Mel-frequency cepstral coefficients (MFCCs). The results indicated that an overly short signal length, such as one second, will weaken the network classification capability (reduce about 2-3% in accuracy). A comparative study was performed on deep convolutional neural networks (CNNs) to assess optimum time-frequency representations (TFRs) as input features for PCG classification. The results showed that continuous wavelet transform (CWT) and chirplet transform (CT) were slightly better than other TFRs including short-time Fourier transfer (STFT), Wigner-Ville distribution (WVD) and Choi-William distribution (CWD). Meanwhile, the appropriate increase of the CNN capacity and architecture optimisation can improve the performance, while the network architecture should not be overly complicated. Using prior knowledge and experience from the mentioned studies, A Hierarchical Multi-Scale Convolutional Network (HMS-Net) was proposed and won the first prize in the CinC/PhysioNet 2022 PCG classification challenge.

3) To design the deep learning algorithm for the detection of paroxysmal atrial fibrillation using single-lead ECG. A two-stage RNN network was proposed during the China Physiological Signal Challenge 2021 (CPSC 2021) which had satisfying performance and held the advantage of low computing load. It showed promising potential for terminal equipment such as the miniaturized ECG-PCG device.

4) To provide accurate respiratory rate while eliminating the need of physical respiration sensor. A study was conducted in the ECG-derived respiration (EDR) field to assess the feasibility of extracting EDR from the localised ECG at the auscultation sites by experiments. Results indicated that the ECG acquisition location barely affected the calculated respiratory rate accuracy. This proved the possibility of providing reliable respiratory rate from the ECG-PCG device without adding an extra sensor. In addition, the study was also conducted to compare the effect of using embroidered and gel electrodes on the extraction of EDR. Despite the slightly poorer performance of embroidered electrodes compared with gel electrodes, embroidered electrodes showed potentials in future low-cost applications. Stress test were also conducted for EDR by experiments. Current results indicated that the artefacts caused by body movement affected greatly on the EDR extraction. Reduce the swing of the ECG wires or using wireless ECG instead may be feasible solutions to improve the performance.

Abbreviations

AM: Amplitude Modulation
ANN: Artificial Neural Network
AV: Atrioventricular
BiLSTM: Bidirectional Long Short-Term Memory
BP: Band-pass
BRNN: Bidirectional Recurrent Neural Network
BW: Baseline Wander
CNN: Convolutional Neural Network
CT: Chirplet Transform
CVD: Cardiovascular Disease
CWD: Choi-Williams Distribution
CWT: Continuous Wavelet Transform
DL: Deep Learning
ECG: Electrocardiogram
EDR: ECG-Derived Respiration
EEG: Electroencephalography
EIT: Electrical Impedance Tomography
EMG: Electromyography
FM: Frequency Modulation
FN: False negative
FP: False positive
GRU: Gated Recurrent Unit
HMM: Hidden Markov Model
HMS-Net: Hierarchical Multi-Scale Convolutional Neural Network
IED: Inter-Electrode Distance
IIR: Infinite Impulse Response
ISOM: Isolated Self-Organizing Map
KNN: K-Nearest Neighbour
LBBB: Left Bundle Branch Block
LMIC: Low and Middle-Income Countries
LSTM: Long Short-Term Memory
MAE: Mean Absolute Errors
MFC: Mel-Frequency Cepstrum
MFCC: Mel-Frequency Cepstrum Coefficient
ML: Machine Learning
MLP: Multi-Layer Perceptron

NAS: Neural Architecture Searching
NN: Neural Network
PCG: Phonocardiogram
RBBB: Right Bundle Branch Block
RR: Respiratory Rate
RSA: Respiratory Sinus Arrhythmia
RNN: Recurrent Neural Network
SA: Sinoatrial
SE: Squeeze-and-Excitation
SHD: Structural Heart Disease
STE: Short-Term Energy
STFT: Short-Time Fourier Transform
SVM: Support Vector Machine
TFD: Time-Frequency Distribution
TFR: Time-Frequency Representation
TP: True Positive
TN: True Negative
VCG: Vectorcardiography
VSD: Ventricular Septal Defect
WHO: World Health Organisation
WVD: Wigner-Ville Distribution

Author's Publication

Thesis Related Publications

[1] **Bao, X.**, Deng, Y., Gall, N. and Kamavuako, E.N., 2020, February. Analysis of ECG and PCG Time Delay around Auscultation Sites. In BIOSIGNALS (pp. 206-213).

[2] **Bao, X.**, Howard, M., Niazi, I.K. and Kamavuako, E.N., 2020, July. Comparison between Embroidered and Gel Electrodes on ECG-Derived Respiration Rate. In 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 2622-2625). IEEE.

[3] **Bao, X.**, Abdala, A.K. and Kamavuako, E.N., 2020. Estimation of the Respiratory Rate from Localised ECG at Different Auscultation Sites. Sensors, 21(1), p.78.

[4] **Bao, X.**, Xu, Y. and Kamavuako, E.N., 2022. The Effect of Signal Duration on the Classification of Heart Sounds: A Deep Learning Approach. Sensors, 22(6), p.2261.

[5] **Bao, X.**, Hu, F., Xu, Y., Trabelsi, M. and Kamavuako, E.N., 2022, February. Paroxysmal Atrial Fibrillation Detection by Combined Recurrent Neural Network and Feature Extraction on ECG Signals. In BIOSIGNALS (pp. 85-90).

[6] **Bao, X.**, Xu, Y., Lam, H.K., Trabelsi, M., Chihi, I., Sidhom, L. and Kamavuako, E.N., 2022. Time-Frequency Distributions of Heart Sound Signals: A Comparative Study using Convolutional Neural Networks. arXiv preprint arXiv:2208.03128. (Under Review)

[7] Xu, Y., **Bao, X.**, Lam, H.K. and Kamavuako, E., 2022. Hierarchical Multi-Scale Convolutional Network for Murmurs Detection on PCG Signals. In Computing in Cardiology (CinC) 2022.

Other Publications during PhD

[8] Kamavuako, E.N., Brown, M., **Bao, X.**, Chihi, I., Pitou, S. and Howard, M., 2021. Affordable Embroidered EMG Electrodes for Myoelectric Control of Prostheses: A Pilot Study. Sensors, 21(15), p.5245.

[9] Wang, B., Hargrove, L., **Bao, X**. and Kamavuako, E.N., 2022. Surface EMG Statistical and Performance Analysis of Targeted-Muscle-Reinnervated (TMR) Transhumeral Prosthesis Users in Home and Laboratory Settings. Sensors, 22(24), p.9849.

[10] Xu, Y., Lam, H.K., Jia, G., Jiang, J., Liao, J. and **Bao, X.**, 2022. Improving COVID-19 CT classification of CNNs by learning parameter-efficient representation. Computers in Biology and Medicine, p.106417.

[11] Zhdanov, A.E., Evdochim, L., **Bao, X.**, Dolganov, A.Y., Borisov, V.I., Kazajkin, V.N., Ponomarev, V.O., Lizunov, A.V. and Dorosinskiy, L.G., 2020, October. Oculusgraphy: Ocular examination for toxicity evaluation based on biomedical signals. In 2020 International Conference on e-Health and Bioengineering (EHB) (pp. 1-6). IEEE.

[12] Zhdanov, A.E., Dolganov, A.Y., Kazajkin, V.N., Ponomarev, V.O., Lizunov, A.V., Borisov, V.I., Lucian, E., **Bao, X.** and Dorosinskiy, L.G., 2020, October. OculusGraphy: Literature review on electrophysiological research methods in ophthalmology and electroretinograms processing using wavelet transform. In 2020 International Conference on e-Health and Bioengineering (EHB) (pp. 1-6). IEEE.

[13] Zhdanov, A.E., Borisov, V.I., Dolganov, A.Y., Lucian, E., **Bao, X.** and Kazaijkin, V.N., 2021, June. OculusGraphy: Filtering of electroretinography response in adults. In 2021 IEEE 22nd International Conference of Young Professionals in Electron Devices and Materials (EDM) (pp. 395-398). IEEE.

[14] Zhdanov, A.E., Borisov, V.I., Dolganov, A.Y., Lucian, E., **Bao, X.** and Kazaijkin, V.N., 2021, June. OculusGraphy: Norms for electroretinogram signals. In 2021 IEEE 22nd International Conference of Young Professionals in Electron Devices and Materials (EDM) (pp. 399-402). IEEE.

[15] Zhdanov, A.E., Borisov, V.I., Lucian, E., Kazaijkin, V.N., **Bao, X.**, Ponomarev, V.O., Dolganov, A.Y. and Lizunov, A.V., 2021, September. OculusGraphy: Description of Electroretinograms Database. In 2021 Third International Conference Neurotechnologies and Neurointerfaces (CNN) (pp. 132-135). IEEE.

[16] Zhdanov, A.E., Borisov, V.I., Lucian, E., Kazaijkin, V.N., **Bao, X.**, Ponomarev, V.O., Dolganov, A.Y. and Lizunov, A.V., 2021, September. OculusGraphy: Filtering of Pediatric Electroretinogram. In 2021 Third International Conference Neurotechnologies and Neurointerfaces (CNN) (pp. 128-131). IEEE.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Background and Motivation

Cardiovascular diseases (CVDs) are the leading cause of death worldwide, which accounts for approximately 31% of the mortality [1]. According to World Health Organization's (WHO's) report, the annual number of death in 2016 was 17.9 million, including around 3.1 million in China, 0.85 million in the USA, 3.9 million in Europe and 2.8 million in India, etc [2]. If the trend continues, the number of mortality is probably estimated to 22.2 million and the global cost on CVDs rises to 1,044 billion dollars by 2030. Currently, there are still over 400 million people living with CVDs, this greatly affects the national productivity and cause economic burden [3]. Besides, it is noteworthy that over three quarters of death occur in the low and middle-income countries (LMICs) where the health systems are weak and lack of standardized protocol-based management, essential medicines, monitoring equipment, and medical staffs [4]. In LMICs, the CVDs risk assessments are not offered in many primary care facilities, so patients may miss the best time for diagnosis [2]. This directly increases the difficulty and cost of treatment. Therefore, WHO proposed 'HEARTS technical package' in 2016, which highlights the importance of primary health care in CVDs prevention [5]. In the high income countries, though the medical condition is much better, it is still a huge expense in the primary care. According to 'European Cardiovascular Disease Statistics 2017 edition', the cost of primary care in the EU is approximately 9.5 billion euro in 2015. In the UK, it is around 1.65 billion euro [6]. Therefore, more automatic, precise and cheaper pre-diagnosis device or system can be of great help to improve the detection onset of heart conditions and reduce the cost in primary care. PCG and ECG are the two major signals of interest for the CVDs diagnosis. This thesis focuses on using bio-medical signal processing and machine learning techniques to improve the two cardiac signal monitoring and computer-aided diagnosis for the future low-cost and portable devices.

1.2 Primary Signals

1.2.1 Phonocardiogram

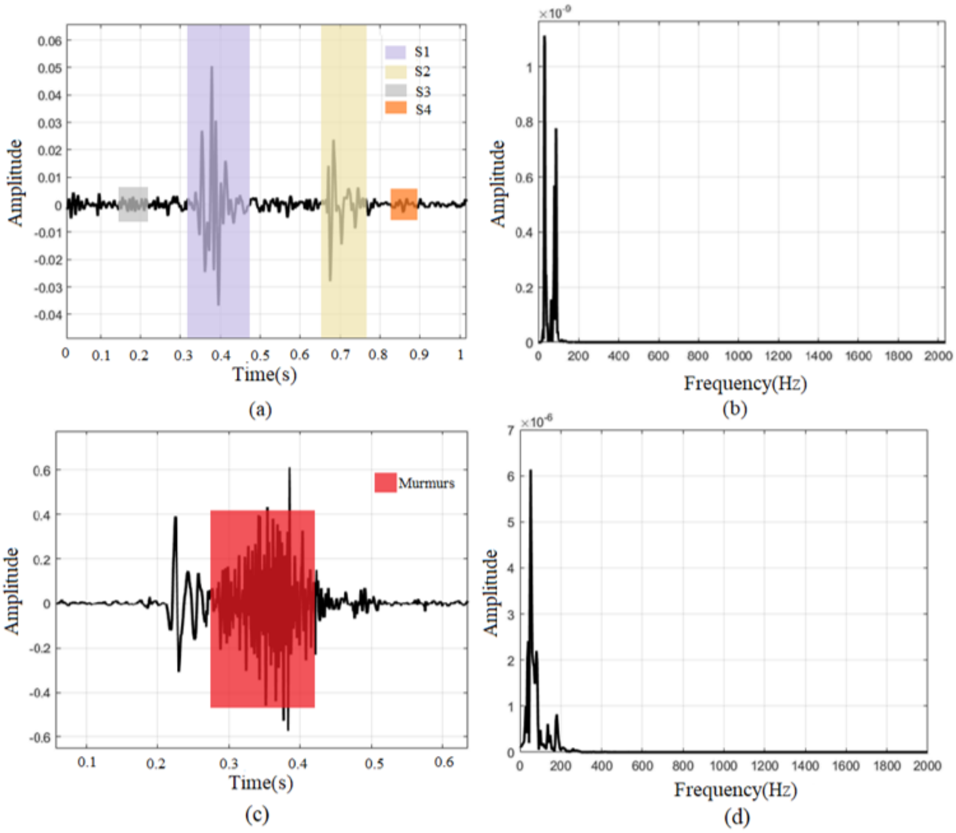
Auscultation is the most common and effective way in early screening, which plays a vital role in CVD detection for necessary action to lower the risk of worsening heart diseases. The physicians use a stethoscope to listen to the murmurs in heart sound at different auscultation sites of the patients and make pre-diagnosis. Heart sound is produced by the closure of a valve or tensing of a chordae tendineae, and the visualized signal of heart sound is PCG as shown in Figure 1.1(a). In healthy conditions, the first heart sound (S1) and second heart sound (S2) are the two main components in the waveform, representing the sound of mitral and tricuspid valve closure (S1) and the closure of the aortic and pulmonic valves (S2), respectively. In addition, the third heart sound (S3) and fourth heart sound (S4) are also innocent components seen on children's PCG but rarely seen in adults. They indicate the sound caused by an increase in ventricular blood volume and an atrial gallop by blood being forced into a stiff ventricle. Figure 1.1(b) shows the full-band frequency of this heart sound. The properties of the normal heart sound components are summarized in Table 1.1.

Table 1.1 Heart Sound Components and their properties.

Heart Sound	S1	S2	S3	S4
Duration (ms)	100–160	80–140	40–50	40–50
Frequency (Hz)	30–50	40–70	<30	<20
Occurrence	Sound of mitral and tricuspid valve closure	Sound of aortic and pulmonic valve closure	The sound caused by an increase in ventricular blood volume	Sound of an atrial gallop produced by blood being forced into a stiff ventricle

Pathological heart sounds differ from healthy ones due to the murmurs primarily caused by the abnormal heart structure, such as turbulent blood flow caused by narrowed or leaking valves [7]. Murmurs can occur in the systolic interval or the diastolic interval, reflecting different types of CVDs, e.g., the most common mitral or aortic stenosis murmurs can be seen during systole as shown in Figure 1.1(c) or Figure 1.2 (a) [8]. Figure 1.2 includes some common types of murmurs with their spectrums. The frequency band of murmurs can be ranged from 20-1000 Hz which is overlapped with heart sound, lung sounds (50-1000 Hz) and bowel sounds (150-2000 Hz) [9, 10]. In addition, there are usually ambient noise and artifacts such as friction and body movement mingled in the PCG recording. Therefore, it is challenging to identify the murmurs especially when the heart lesion is in early stage.

As shown in Table 1.1, the heart sound components are short in time and low in frequency, and its principal frequencies are at the lower end of the human ear (20–20k Hz),



(a) (b) (c) (d)

Fig. 1.1 (a) Visualisation of a healthy heart sound signal with its component locations. (b) Spectrum of the healthy heart sound signal. (c) Murmurs in the heart sound. (d) Spectrum of the heart sound with murmurs.

which is not sensitive to hear. Therefore, traditional auscultation has natural shortcomings. Furthermore, in the heart sound analysis, murmurs are not always clearly visible, this also causes misdiagnosis. As a result, it relies significantly on the physician's listening ability and clinical experience. Experienced cardiologists can distinguish between 73% and 80% accuracy of pathological murmurs, while inexperienced new physicians or trainees can distinguish 20–40% accuracy [11, 12]. A misdiagnosis leads the patient to miss the best time for treatment or increase the cost due to unnecessary further detection (e.g., electrocardiogram, cardiac ultrasound, and computerised tomography).

Computer-aided heart sound analysis has great potential to improve auscultation accuracy by overcoming human hearing limitations and clinical experience. However, there are still a lot of challenges to be addressed before practical use. For example, not robust heart sound segmentation techniques, unsatisfying classification accuracy in practical use or imbalanced datasets for training machine learning models due to the limitation of data acquisition. The more detailed review and research gap are presented in Section 2.1.

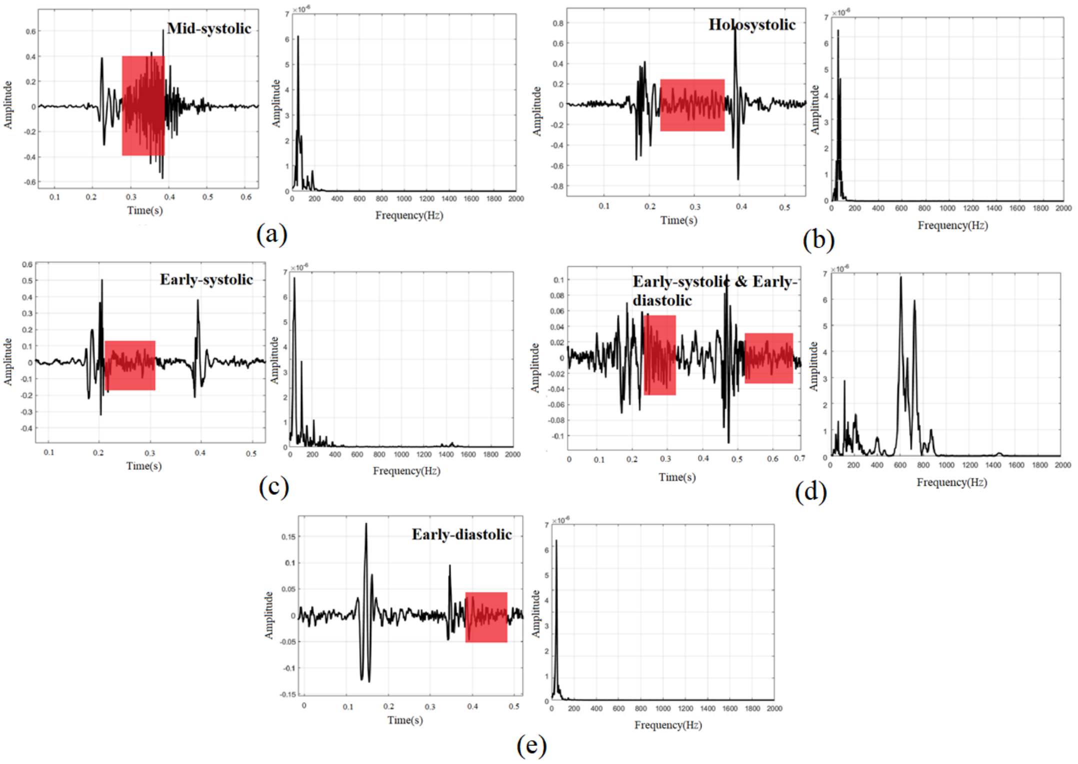


Fig. 1.2 (a) Mid-systolic murmur. (b) Holosystolic murmur. (c) Early-systolic murmur. (d) Early-systolic & early diastolic murmurs. (e) Ealy-diastolic murmur.

1.2.2 Electrocardiogram

ECG is another commonly used approach in cardiac monitoring and diagnosis. It represents the electrical activity of the heart and its waveform is as shown in Figure 1.3. The whole electrical process starts with the spontaneous impulse generated at the Sinoatrial node (SA node), then propagates to the atrioventricular node (AV node), causing the squeezing of the atria as represented by the P wave. Afterwards, the electrical signal is transmitted to the His bundle and Purkinje fibres, causing the contraction of the ventricles. The ventricles are repolarized and ready for the next heart cycle. The QRS complex indicates the depolarization, and the T wave shows the repolarization of the ventricles, respectively. Thus duration and amplitude of these waves provide significant information for diagnosis of heart conditions.

In clinical diagnosis, 12-leads ECG is used to provide more information of the electrical propagation from a larger surface area surrounding the heart. With the ECG signals from different angles and directions, the cardiologists can identify and locate the pathological changes on the heart, such as the bundle branch blocks, localised conduction disorder, premature beats, bradycardia, tachycardia, etc. In the portable and miniaturized devices, the clinically used 12-lead ECG is no longer suitable, reduced or one lead ECG can

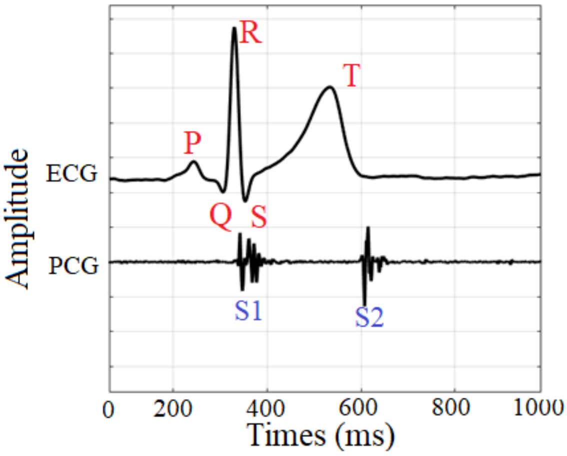


Fig. 1.3 A complete ECG cycle along with PCG.

still provide accurate heart rate to check various arrhythmias and can be used for PCG segmentation according to the time relationship as shown in Figure 1.3. In general, the S1 onset is with QRS complex which are related with the procedure of ventricles contraction and the closure of the bicuspid (mitral) and tricuspid value after the contraction. S2 follows T wave which are related with the procedure of blood pumping out of the heart by aortic and pulmonic valves while ventricles repolarize for next heart cycle.

In this project, ECG is mainly studied for heart rate monitoring and PCG segmentation on a miniaturized portable device. Furthermore, the diagnostic capabilities of single lead ECG signal is explored using machine learning techniques. The detailed research motivations and objectives on ECG are given in Section 2.2.

1.2.3 Respiratory Rate

Respiratory rate (RR) is another physiological indicator focused in this project. It indicates the breaths per minute, which is commonly used as an early warning sign in disease detection. The normal RR of a healthy adult at rest is between 12–16 bpm [13]. Compared with adults, children's RR is higher. For an infant, it ranges from 30–60 bpm, and with growth, the RR gradually reaches the adult level [14]. The resting RR of older people may slightly increase. For the healthy independent seniors, it is 12–20 bpm, and those who need long-term care can reach 16–25 bpm [15]. Generally, a resting RR outside of these ranges may indicate a potential disease. An increased RR (tachypnea) may suggest fever, dehydration, asthma, chronic obstructive pulmonary disease, heart disease, etc. [16]. A low RR (bradypnea) may reveal the use of narcotics, alcohol intake, abnormal body metabolism, sleep apnoea, etc.

Usually, the RR is calculated by counting the number of chest undulations in one minute outside of intensive care [17]. However, this method is time-consuming and cannot do continuous monitoring. In addition to manual counting, there are also devices for measuring RR, such as the widely used chest strap or impedance-based pneumography device [18]. Though, wearing a chest strap causes a sense of pressure when breathing, so it is not suitable for long-term monitoring. Therefore, a portable and wearable method of automatic RR measurement is in great need to effectively monitor breathing in real-time and detect the first sign of physical deterioration in a timely manner. As part of the cardiac monitoring research, studies can be conducted on using bio-signal processing techniques to acquire RR without extra sensors, such as extracting from ECG and PCG. This can contribute to design small-scale ECG-PCG integrated device and providing accurate RR for monitoring or assessment. In the meantime, it can be of great commercial value for the clinic and other researchers for its capability to measure multiple physiological parameters effortlessly.

1.3 Thesis Organization

This thesis includes eight chapters in total, of which Chapter 4, 5, 6 and 7 are the main technical chapters involving the conducted research. In this chapter, the background, motivation, and epidemiology of this project were discussed. Furthermore, the mainly focused physiological signals including PCG, ECG, and respiration were introduced. Their potential research directions were also indicated including the PCG and ECG classification, ECG for PCG segmentation, and respiratory signals extraction from ECG.

Chapter 2 provides a more detailed and comprehensive literature review of the above potential research directions to clarify the research state of the art, and provide a theoretical basis for the proposed research objectives in Chapter 3.

Chapter 3 is concise that highlights the research questions in the project and the research strategy to answer them. This aims to facilitate readers to quickly understand and read.

Chapter 4 presents the analysis of the temporal relationship between ECG and PCG signals at the auscultation sites. The purpose of this study is for a miniaturized ECG-PCG integrated device to simultaneously provide two signal measurements while performing accurate and robust PCG segmentation using the auscultation sites localised ECG as reference.

Chapter 5 is the most informative chapter of this paper. It discusses the input optimization of deep learning networks to improve the accuracy of PCG signal classification. The effect of input signal length on different RNN networks are analysed, and the performance of different time-frequency representations (TFRs) as CNN input are compared. Using the prior knowledge gained in the conducted research, the proposed winning algorithm in PhysioNet 2022 PCG Classification Challenge is presented with its result.

Chapter 6 includes the study of ECG classification using deep learning. It presents the proposed two-stage RNN for paroxysmal atrial fibrillation detection during China Physiological Signal Challenge 2021.

Chapter 7 describes the research on EDR technology. The aim is for the future miniaturized ECG-PCG integrated devices to provide reliable respiration rate while eliminating the necessary of extra respiratory sensors. Therefore, the effects of using auscultation site localised ECG to extract respiratory signals and the feasibility of using embroidered electrodes were studied.

Chapter 8 concludes the thesis and discusses the possible research questions for future work.

Chapter 2

Literature Review

Following the focused signals mentioned in Section 1.2, this section illustrates the review outcome in the PCG and ECG classification, including the analysis system and existing classification methods. For Respiratory Rate, the review is conducted on the extracting approaches from synchronous ECG signals. By reviewing these information, the research gaps and objectives is concluded in Chapter 3.

2.1 PCG Signal Classification

Computer-aided PCG analysis is of immense importance and potential to improve auscultation accuracy. It can overcome the clinical experience and listening limitation by using machine learning algorithms to learn the diagnostic ability on large amounts of data. A typical PCG signal classification procedure is as shown in Figure 2.1, including filtration, segmentation, feature extraction and classification.

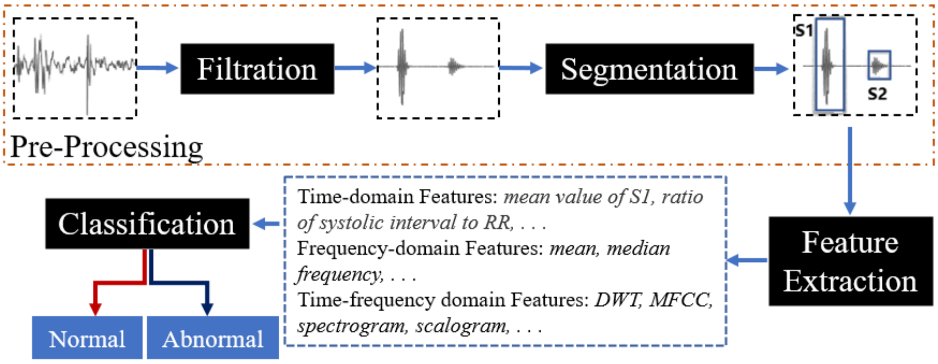


Fig. 2.1 Typical steps for PCG signal classification.

2.1.1 Filtration

In the recording of biomedical signals including PCG signals, the ambient noise and artifacts such as friction and body movement are unavoidable. Besides the external noises, the internal noises from the device circuit also exist, such as the white noise and powerline interference. Therefore, the filtration is needed to reduce these noises. In general, the filtration methods can be basically divided into: (1) linear filters [19–22] and (2) wavelets [23–28]. Actually, there is no widely recognised better method or settings for filtration. Because the selection of filtering method and the filter settings should depend on the acquisition settings and quality of the data. However, it is worth noting that the premise of filtration is avoiding signal distortion. Over-filtered signals may lose the diagnostic information. For instance, the lung sounds and bowel sounds are noises coupled with several types of murmurs in the frequency bands. As a result, these noises cannot be eliminated completely and separately while the murmurs retained. On the contrary, under-filtered (lower order or broader frequency band) can barely make efforts on the subsequent processing.

2.1.2 Segmentation

In standard signal processing, segmentation means cutting the signal into segments by moving windows. In contrast, heart sound processing means breaking the signal into heart cycles and indicating the heart sound components. To avoid confusion, the ‘segmentation’ in this thesis means the conventional heart sound segmentation, and other verbs are used to describe signal sliding.

During segmentation, the heart sound components including S1, S2, systole and diastole can be located from the PCG signals. The success of segmentation can directly affect the extraction of common features, such as the ratio of systolic to diastolic interval in the time-domain, median frequency of the systole in the frequency-domain. Subsequently, the final classification can not be accurate with failed segmentation. All the proposed segmentation methods can be basically grouped into: (1) ECG reference based methods use the R-peak and T wave to determine the locations of the heart sounds. It highly requires the simultaneous recording of the ECG and PCG signals [29, 30], but robust in performance and computationally efficient; (2) Envelope-based methods are more commonly used techniques in non-ECG based segmentation. They use the signal energy to do morphological transformation [31–33], but their performance decrease in the presence of large environmental noise and murmurs; (3) Temporal-spectral parameters based methods use the time-frequency domain characteristics of the heart sounds, murmurs, and noise to segment the heart sound [34, 35]. Same as envelope-based methods, their performances on signals with murmurs and noises are not satisfying; (4) Wavelet based PCG segmentation methods are the evolution of temporal-spectral parameters based methods [36–39]. They decompose the signals to emphasize the heart sounds and suppress the effects of murmurs and noises. The major challenge of wavelet-based segmentation method is to select the appropriate filters, decomposition level and required sub-bands for heart sound and murmurs detection; (5) Hidden Markov models (HMM) are also used for segmentation in recent years [40, 41, 22], and they have outstanding performance in low signal-to-noise ratio. At present, there is no widely recognized sufficiently effective PCG segmentation method, but with the presence of simultaneous ECG recording, ECG-based segmentation is more desirable for practical applications due to its robustness and simplicity.

2.1.3 Feature Extraction

Feature extraction is an important stage in any classification work, which can directly affect the performance. For PCG classification, there are various features that can be selected. For instance, from the morphology of the signal, features related with the component amplitude and appearance can be extracted in the time domain, such as the mean and standard deviation of S1, S2, systole and diastolic interval durations, the ratio of single beat's systole and diastole durations, and the ratio of the mean absolute amplitude during systole to that during the S1 period in each heart beat [42, 43]. After converting the signal into the frequency domain by Fourier transform, its frequency information such as max, mean or median frequency value of the components, discrete energy spectrum, spectral energy in specific frequency bands can also be used as features [44–47]. It is worth noting that the extraction of these time domain or frequency domain features greatly rely on the successful segmentation. In the time-frequency domain, both time and frequency information are provided at the same time which benefits a wide range of signal analysis, such as optics, acoustics, and biomedicine [48]. There are various methods that can convert the signal from 1-D time domain into 2-D time-frequency domain, including short-time Fourier transform (STFT), continuous wavelet transform (CWT), Chirplet transform (CT) and Wigner-Ville distribution (WVD). The acquired time-frequency representations (TFRs) after conversion are used as features for heart sound classification [49–51]. However, due to the large size of TFRs as input, they are normally used in deep learning, such as Convolutional Neural Networks (CNNs). Therefore, mel-frequency cepstrum coefficients (MFCC) and wavelet are more commonly used features [28, 49, 52–56]. They can be somehow understood as simplifying the frequency domain information and only retaining the effective coefficients among the frequency bands. As a result, they can hold both information from time and frequency domain while retain effective features. In general, there is no widely-recognized best feature for heart sound classification, and the features are sometimes used in combination. The selection of the features depends more on the classification method.

2.1.4 Classification

Classification is to use the machine learning methods to classify the input patterns into specific classes. For PCG classification, it is the finial step to decide the health condition of the heart based on the extracted features. In general, the mainstreamed classification methods can be divided into two types: (1) traditional machine learning, including logistic regression [57], random forest [58], K-nearest neighbours (KNN) [49, 52, 57, 59], regression tree [49], isolated self-organizing map (ISOM) [53], support vector machine (SVM) [52, 55, 57, 60–62] and HMM [63]; (2) neural network based approaches, such as artificial neural network (ANN) and its variants [28, 50, 51, 54, 64–70]. In recent years, with the development of neural networks and deep learning, they are more used for PCG classification. Compared with traditional machine learning, they can weaken the dependence on segmentation and feature extraction, thereby enhancing the robustness of the algorithm. Moreover, they also performed promising classification in terms of accuracy. Therefore, an important research question of this thesis is to explore how to optimize the input of deep learning to improve the classification accuracy.

2.2 ECG Signal Classification

ECG signal contains important diagnostic information on cardiovascular disease and cardiac arrhythmia, therefore, accurate recognition and classification of ECG signal is a hot topic in the biomedical research. Nonetheless, there are still remaining problems to be addressed from practical use in the automatic classification. Because the morphological and temporal characteristics of the ECG signals are significantly different from patient to patient and varies by physical conditions[71–73]. For instance, sometimes the morphologies are different for the same disease between patients or two different diseases show the same characteristics on ECG signals.

Furthermore, compared with PCG, ECG signal has stronger monitoring significance. It is often used for long-term detection such as bedside monitoring or wearable online health monitoring. However, in the detection of abnormal heartbeats, each heart cycle in the ECG needs to be analysed, which can be cumbersome and time-consuming for human interpretation. This loses the timeliness of monitoring. Besides, the human interpretation also requires deep knowledge in the cardiology and clinical experience. Therefore, reliable computer-aided algorithms are needed to improve the situation.

The typical ECG signal classification procedure is similar to PCG classification, including pre-processing, feature extraction and classification.

2.2.1 Pre-Processing

In the ECG recordings, there are always artifacts mingled with the raw signals. Therefore, the first step is to reduce the noise. The commonly existing noise include: (1) Powerline interference, which is 50 (or 60 Hz) within 1 Hz frequency band. (2) Baseline wander, which is normally caused by the breathing induced chest movement. Its frequency is between 0.15 to 0.3 Hz. (3) Electrode contact noise, which is the result of deficiency electrode-skin contact and unstable impedance. This can usually result in complete distortion of the signal. (4) Muscular interference, which is generated by the muscle contraction apart from the heart. The frequency band is between 0-500 Hz and dominant in 50-150 Hz. (5) Instrumentation noise, which is produced measuring device internal circuit. In the filtering stage, a band-pass filter is applied. The recommended frequency range is 0.5-40 Hz [74], but other frequency bands such as 1-40Hz [75], 1-30Hz [76], and 0.1-100Hz [77] were also used in the previous studies.

After ECG filtering, resampling and normalization are two optional steps depending on the data sets. Resampling aims to reduce the computational cost, and for ECG, the signals can be downsampled to as low as 100 Hz without loss of useful information. Other downsampling frequency such as 125 Hz [78], 250 Hz [79] or 360 Hz [80] were also applied. Normalization can avoid amplitude biases between signals and DC offset effects.

2.2.2 Feature Extraction

R peak detection is the first and vital step in the feature extraction. The located R peaks can provide the heart rate which is a important indicator for arrhythmia screening, and can also help in the further detection of other ECG components. The most primitive method of is using past knowledge to set threshold on ECG amplitude to detect R-peaks, which is obviously not accurate and robust enough when facing noises, pathiological signals and specific ECG leads. Afterwards, methods include wavelet transform based methods[81–86], mathematical morphology based methods [87, 88], digital filtering methods [89–91], and Shannon energy envelope-based methods were proposed [81, 92–95]. The Pan-Tompkins method is the most well-known benchmark which is the combination of band-pass filter, derivative filter and squaring to detect the R peaks [89].

In literature, various features has been proposed according to different analysis or classification aims. Overall, these features can be divided into: (1) morphological features, most are based on QRS, including R wave duration, R peak amplitude, QRS wave area, QS amplitude difference, QR slope, etc. [81, 96–104]; (2) statistical features, such as maximum, minimum, mean, standard deviation, energy, entropy, kurtosis and skewness [72, 105–110]. They are effective in analyzing the level of complexity and the distribution in the time-series; (3) wavelet based features, which are powerful in the extraction of time-frequency localization property [98, 111–115]. The advantage of the wavelet is preserving both approximations and details during full sub-band decomposition which can benefit the analysis; 4) other features, such as power spectral density-based attributes [112, 113, 116, 117], auto-correlation analysis-based features [75, 118] and TFRs [119–122] were also applied in the previous research. These features can be used independently or in combination according to the detection purposes.

2.2.3 Classification

The analysis and classification of ECG can be used for disease diagnosis, heartbeat type detection, biometric identification, emotion recognition, etc. The review in this part mainly focused on the first two aims. In general, the classification methods used for ECG are basically the same as that of PCG. The main differences are the input features. Therefore, mainstreamed ECG classification methods can also be divided into: (1) traditional machine learning, including linear discriminant analysis [79, 99, 109], KNN [79, 109, 123, 124], decision tree [79, 125–127], SVM [128–131], bayesian [126, 129], linear regression [132], logistic regression [109, 125, 133], etc; (2) neural network based approaches, including CNN [119–122], recurrent neural network (RNN) [116, 134–136], fuzzy clustering ANN [137–139], time-delayed neural network [140], generalized regression neural network [141], etc. Since most of the previous studies were with different specific research aims and applied different databases, it is impossible to objectively compare the performance of these classifiers only by review. However, deep learning methods have a broader research space due to their powerful feature mining ability that even raw ECG signals can be used as input. Therefore, this thesis also focuses on the use of deep learning for one lead ECG recognition.

2.3 ECG Derived Respiratory Rate (EDR)

In critical care (or intensive care unit, ICU), RR is also a vital parameter in the monitoring of respiratory failure. It could be measured by the gas exchange using a ventilator, capnography monitors, or spirometry devices, and chest electrical activities using electrical impedance tomography (EIT), inductance plethysmography, or impedance pneumography [142, 143].

2.3.1 Respiration Rate Measurements

The current measurement of RR outside of the critical care still relies on manually counting the chest undulations in one minute by the medical staff [144]. Although this practice is easy to conduct without using extra medical devices, it has some drawbacks such as (1) low accuracy. Subject's awareness, poor visibility of a breath, and other interruptions which greatly affect the measurement. Besides, in practice, the manual counting is not completed in one full minute by the medical staff due to the heavy workloads. They usually multiply the 30 s or 15 s measurement by 2 or 4 to assess the RR, which leads to further inaccuracies [145]; (2) it is labour-consuming, as the medical staff can only conduct the measurement on one patient at one time; (3) the measurement is not continuous. As an early sign of physical deterioration, real-time and continuous monitoring can help alert the staff to emergencies, such as heart failure, shock, diabetic coma. However, the intermittent measurement cannot provide such information timely, so the RR is always underutilized. External devices to automate the RR measurement can remedy the deficiencies associated with manual counting to a certain extent. Despite this, there are still respective limitations to each method. For the gas exchange-based techniques, they are accurate methods to reflect the respiratory condition, but have no portability, which requires the patients breathing in the external tube of the devices. So, these techniques are generally only available in critical care [18]. The bioimpedance-based techniques such as impedance pneumography can measure the electrical activities on the chest during inhalation and exhalation. However, it requires the patients to wear a tight chest strap, which may cause discomfort [146]. Additionally, patient movement, bad contact, and obstruction of breath can cause inaccurate measurements. Acoustic sensors are also used in the measurement of RR, however, their performance is affected by the environmental noise and skin friction [147]. Therefore, wearable devices for automatic RR measurement are in great need to effectively monitor the breath in real-time and detect the first sign of physical deterioration promptly.

2.3.2 EDR Principle and Extraction Methods

Extracting respiratory signals from the ECG signals is a potential surrogate measurement of RR. In recent years, ECG devices are becoming miniaturized, and sensors have been integrated with sport bands, smartwatches, and other portable monitors. This provides the feasibility and potentiality to design wearable ECG-based RR measurement devices. The first study on respiration-induced ECG variation was proposed by Einthoven et al. [147]. Flaherty and Riekkinen further analysed the respiration influence on children and cardiac patients by isopotential surface-mapping and vectorcardiography (VCG) [148, 149]. Nowadays, it is well known that respiration-induced ECG variations are caused by (1)

Respiratory Sinus Arrhythmia (RSA) that refers to the cyclic variation that the heart rate accelerates during inhalation and decelerates during exhalation [150]. It can be reflected in the ECG signals as the frequency modulation (FM) of the R-R interval between the R peaks as shown in 2.2 (a). (2) Respiration-induced electrical axis rotation. During the inspiration, the filling of the lungs stretches the heart apex towards the abdomen, and in expiration, the emptying of the lungs compresses the heart towards the breast. Due to the displacement of the heart, the electric cardiac vector will change during respiration [151]. In the ECG signal, this process can be indicated as amplitude modulation (AM) of the R peaks as shown in 2.2 (b). (3) Baseline Wander (BW) is the artefact caused by body movement, including breathing. The expansion and contraction of the thoracic cavity due to respiration cause a slow and undulating baseline in the ECG signals as depicted in 2.2 (c) [152].

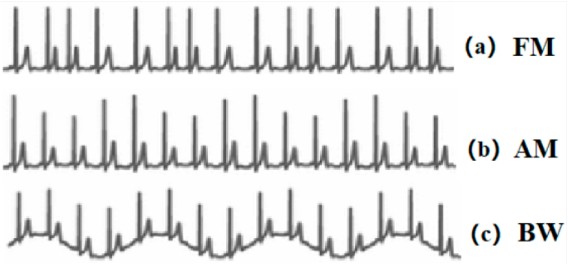


Fig. 2.2 (a) Respiratory Sinus Arrhythmia (RSA) induced frequency modulation (FM). (b) Electrical axis rotation caused amplitude modulation (AM). (c) Baseline wander (BW) caused by chest movement.

Several techniques to extract respiratory signals from the ECG, the so-called ECG-derived respiration (EDR), have been proposed according to the respiration-induced ECG variation mentioned above. Some techniques are based on multi-leads ECG signals [151–155], while others attempt to extract respiratory information from one-lead ECG [156–161], as well as direct band-pass (BP) filtering of the ECG within the respiratory frequency band [152, 159, 162]. For the multi-leads EDR techniques, they mainly use the rotation angles of VCG from multiple ECG leads, while the one-lead EDR methods focus on the features related to the QRS complex, such as amplitude, interval, area, slopes. There is no consensus on which one is better in the performance; however, for a wearable device, one-lead ECG has the advantage in the system complexity and size. Since this thesis aims at the future wearable ECG-PCG integrated device, the current EDR research is mainly focused on standard ECG leads, especially lead II which is poor on portability. Therefore, it is necessary to investigate whether measuring ECGs at unconventional locations, such as chest with shortened electrode distances affects the EDR extraction.

2.4 Review Outcome

In Section 2.1 and 2.2, the steps of PCG and ECG classification and related approaches are introduced in detail. For PCG classification, segmentation is an important step in traditional machine learning methods. It directly affects the performance of signal feature extraction and classification. PCG segmentation is still an open question in the heart sound analysis. ECG can be a reliable and stable reference for PCG segmentation. When there are murmurs or noise underlying within the PCG signals, signal processing methods for segmentation are normally not robust enough. As a result, it may affect the feature extraction and classification performance. Therefore, integrating ECG with the system can be an effective approach to solve this issue. However, it is unclear how is the auscultation localized ECG like and there is no existing database to analyze. Considering that there is electrical axis deflection when ECG signals are collected at different body positions, and there may also exist time delay in the propagation of heart sounds on the chest. Therefore, in the future miniaturized PCG-ECG integrated device, understanding and verifying the time relationship of the two signals is the prerequisite to ECG for PCG segmentation.

In terms of PCG classification algorithms, the research has been conducted for decades. Many machine learning approaches have been proposed, however, due to accuracy, portability and robustness issues, these is still distance from practical deployment of these algorithms. In recent years, with the development of deep learning and advanced computing hardware, using deep learning to improve the recognition of heart sounds is a promising solution. At present, deep learning methods mainly use spectrogram by STFT and scalogram by CWT as inputs. It is worth studying to explore more effective input features or optimize inputs to improve the classification performance. In addition, during the review, we also found that the use of neural networks in many studies is not that reasonable. Some networks are too small to extract sufficiently deep features and fully fit the input samples, or excessively use complex neural networks, resulting in huge computing load which does not have the potential for embedded or local device. Therefore, designing more reasonable neural network for PCG classification is also a direction we can study.

Similar issues exist in ECG classification. Compared with PCG, which is mainly aimed at diagnosis, ECG has more monitoring capacities and can be used for long-term detection of cardiac lesions that are not obvious in the early stage. Furthermore, since ECG can reflect a variety of heart conditions, it is necessary to design algorithms for specific heart conditions. Therefore, the research on ECG classification can focus on paroxysmal heart conditions.

2.3 introduced the respiration monitoring methods and the principle of EDR technology. Respiratory rate is a critical indicator in the cardiac monitoring, and it is necessary to be provided by the screening system. To make the device low-cost and miniaturized, it is possible to use EDR techniques rather than extra respiratory sensors to extract the respiratory signal. At present, the main EDR researches are aimed at the conventional 12-lead ECG. For the miniaturized PCG-ECG integrated equipment, the use of localized ECG for EDR has not been studied. So, it is also necessary to verify its feasibility. Moreover, conventional 12-leads ECG signals are collected by gel electrodes, which is not suitable for future portable devices. Therefore, selecting appropriate sensor materials is also a question worth researching. In addition, the researches on EDR are mainly at static, and there are not many studies under the stress test. So, research is also needed.

Chapter 3

Thesis Objectives and Research Strategy

This Chapter briefly summarizes the thesis aim, presents the objectives proposed after the review in Chapter 2, and introduces the opted research strategy to address these objectives.

3.1 Aims and Objectives

As stated previously, early cardiac screening is of significant importance to prevent the lesions deterioration. However, the accuracy of manual auscultation by physicians is not high enough at present [38, 12]. As a result, most of the diagnosis rely on the secondary screening modality such as ultrasound and computed tomography scan. But these detections are not easily accessible, especially for the people in rural place and low- and middle-income countries. In addition, these detections also require professional physician interpretation. For large-scale cardiac screening, these detections not only increase the cost of patients, but also consumes the medical resources of the hospital. Therefore, a low-cost and portable system is particularly needed for improving the early cardiac screening. The system can have trustworthy computer-aided heart sound analysis capabilities, and can provide measurement of various physiological signals and indicators, making early screening more efficient and reliable. This thesis aims at advancing the realization of this system from two aspects : (1) developing machine learning techniques to improve the computer-aided diagnosis of heart sound and ECG; (2) assessing the feasibility by experiments, utilising biomedical signal processing to eliminate the need of physical respiration sensor.

As discussed in Section 2.4, the proposed specific objectives includes:

1) To assess the feasibility of localised ECG signal acquisition and analyse its usability for PCG segmentation.

2) To investigate the optimal use of deep learning input and propose a reliable algorithm for PCG classification.

3) To design the deep learning algorithm for the detection of paroxysmal atrial fibrillation using single-lead ECG.

4) To provide accurate respiratory rate while eliminating the need of physical respiration sensor.

3.2 Overview of Research Strategy

Since there is no existing database of auscultation localized ECG. The first objective can be addressed by experiments to collect our own ECG-PCG data to test the feasibility of measuring localized ECG and analyze the ECG-PCG time relationships to figure out how the localized ECG can be used for PCG segmentation (Study I, Chapter 4).

The acquisition time period of pathological signals is quite long, so it is necessary to use open-sourced database such as PhysioNet which has relatively large amount of PCG data. The second objective is addressed by using these databases or attending world-class challenges in this area (Study II, Study III and Study IV, Chapter 5).

The third objective is addressed same as the second one by attending challenges. This solves the ECG data issue and can insure the research is meaningful (Study V, Chapter 6).

The forth objective is to assess the feasibility of using localized ECG for EDR. Since there is no existing suitable database, this question is addressed by experiments to collect data for analysis. Besides, for recycle use of the ECG sensors, we also explored using embroidered electrodes for EDR. The stress test is also conducted to verify the performance (Study VI, Study VII and Study VIII, Chapter 7).

Chapter 4

Analysis of ECG and PCG Time Delay around Auscultation Sites

4.1 Introduction

Heart sound auscultation and Electrocardiogram (ECG) are the two most common and effective ways in the primary diagnosis of heart diseases. Their signal waveforms, phonocardiogram (PCG) and ECG can reflect the mechanical and electrical activities of the heart, respectively. The PCG signal can reveal the physiological or pathological conditions of cardiac valves and chambers to diagnose the structural heart disease (SHD), such as prolapsed mitral valve, ventricular septal defect (VSD), tricuspid regurgitation. The ECG can help to detect diseases associated with impulse conduction, such as arrhythmias, coronary heart disease, heart attacks, etc. [163].

The normal cardiac cycle relies on the cooperation of electrical activity and mechanical contraction of the atria and ventricles of the heart. The whole process is initially stimulated by the spontaneous action potential in the sinoatrial (SA) node (represented as P wave on ECG), then propagated to the atrioventricular (AV) node causing the atria contraction and the blood is pumped into ventricles and the ventricular depolarization (represented as QRS complex on ECG) begins. Once the ventricular pressure becomes greater than the atrial pressure, the atrioventricular valves close (represented as S1 onset on PCG) and the ventricular depolarization is finished. The continuation of the electrical signal goes through the bundle of His to the Purkinje fibres causing the ventricle contraction and the blood is pumped out of the heart. After the blood pumping out, the ventricles are repolarized (represented as T wave on PCG) and relaxed. The closure of the semilunar valves cause the S2 on the PCG. Therefore, the PCG and ECG are closely related in the time domain [164].

In order to fully utilise the diagnosis power of the PCG, it is of utmost importance to segment S1 from S2. As described in 2.1.2, there is no widely recognized best PCG segmentation method, but with the presence of simultaneous ECG recording, ECG-based segmentation is more desirable for practical applications due to its robustness and simplicity. In the previous ECG-base PCG segmentation studies, S1 onset is conventionally considered to occur after R peak [165–168]. [167] detailedly summarized the time property of heart sounds that S1 starts 10–50 ms after R peak and lasts for 100–160 ms; S2 starts 280–360 ms after R-peak in ECG and lasts for 80–140 ms. For the practical applications, the development of microprocessor in the last two decades has made it possible to make portable devices that can be of great value in primary care. Devices, such as the SensiumVitals® system, Zio patch monitor and CAM patch monitor can collect the ECG using a lightweight patch on the chest region. This provides the possibility to integrate ECG and PCG together around the chest auscultation area, instead of measuring at different places of the body. Integrating PCG and ECG together for concurrent measurement is of great help to increase the portability and reduce the size in designing the small and portable device or systems. Furthermore, combining ECG with PCG can provide more comprehensive heart diagnosis [64, 169, 170]. In such case, the need for sophisticated segmentation can be mitigated by using the ECG as reference signal and segmenting the PCG accordingly.

To the best of our knowledge, research on automatic analysis of PCG is mainly based on single channel signals and the time correlation described is not exhaustive on which lead of ECG and auscultation site were used. However, multiple channels auscultation provides more comprehensive information on the heart conditions. On this basis, there are studies on multi-site PCG recording to visualize the heart related acoustic sounds by cardiac acoustic mapping [171–175]. These studies not only provide a new way to analyse the heart sound, but also illustrate that the heart sound generation and propagation delay in the auscultation area. In addition, the ECG signals have morphological changes due to the electrode placement around the chest. According to [176], the QRS complex shifts due to the electrodes placement. Therefore, it is not known whether the correlation between ECG and PCG remains the same when multiple channel signals are collected from different auscultations sites. In the case of small-scale ECG-PCG device, the recordings of the ECG should occur around the auscultation site. It is therefore of utmost importance to revisit the time properties of ECG and PCG.

The primary aim of this study is to analyse the time delay between ECG and PCG at different auscultation sites (A, P, T, M). The secondary aim is to investigate the changes in the time occurrence of the R-peak in relation to the distance between the recording electrodes. All the findings contribute to design small-scale ECG-PCG integrated device and provide more precise time property for ECG-based PCG segmentation.

4.2 Experimental Settings and Processing Methods

4.2.1 Data Collection

1) Subjects

The experiments were conducted on 12 human subjects with no history of heart diseases (8 male/ 4 female, age range 21–28 years, mean 25.6 years). The procedures were approved by the King's College Research Ethics Committee (Approval No.: LRS-18/19-10673). Subjects gave written informed consent prior to the experimental procedures.

2) Experimental Settings

The proposed experiment requires the simultaneous acquisition of ECG and PCG signals at each auscultation site. A simple block diagram of this hardware system is shown in Figure 4.1. The recording uses the commercial acquisition system (iWorx, model RA 834) as recorder. ECG devices (iWire-BIO4) and digital stethoscopes (ThinkLab One) are connected with the recorder by iWire inputs and DIN8 inputs. The solid gel electrodes (Ambu 0215M) are used as ECG sensors. The sampling frequency was 20 kHz to allow fine resolution around the 0.05 ms. The filter for ECG was 0.05 – 40 Hz [177], and PCGs were recorded with wideband mode (20 – 2000 Hz).

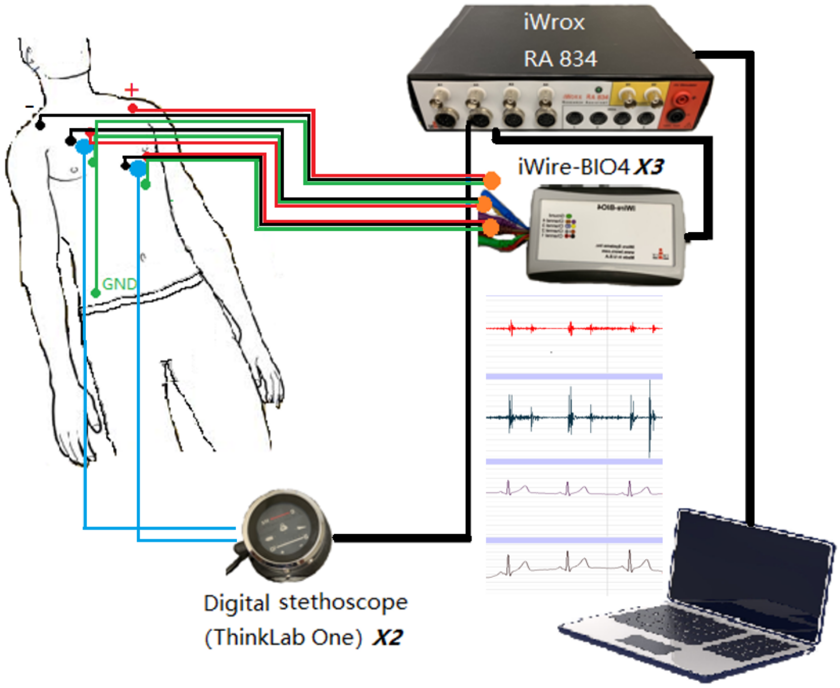


Fig. 4.1 Block diagram of the recording setup.

3) Experimental Procedures

Prior to commencing the experiments, we quantified the instrumentation delay. To achieve this, all three iWire devices where connected on the same electrodes in a limb lead configuration while both microphones where placed close to the auscultation site A. The sampling frequency was set to 100 KHz for this particular measurement. Instrumentation delay was very negligible about 10 microseconds.

The experiments are performed in two stages, and the subjects should keep supine. Stage I is to analyse whether the inter-electrode distance (IED) affects the ECG delays. Three groups of disposable adhesive ECG sensors are placed at A site with 5 cm, 10 cm and 15 cm IED as shown in Figure 4.2(a). The data is collected for 3 mins.

In Stage II, the effect of placement on ECG and PCG delays is analysed. Three electrodes (red points) are positioned over the chest of a subject with standard Lead I as reference. The other two iWire devices do the simultaneous recording with 10 cm IED at each auscultation site. Two ThinkLab stethoscopes are put at the centre of the electrodes (auscultation points). Each site is identified using the anatomical landmark and listening. The placement of sensors in Stage II can be seen in Figure 4.2(b). Four groups of data are collected corresponding to each auscultation site with 3 mins duration.

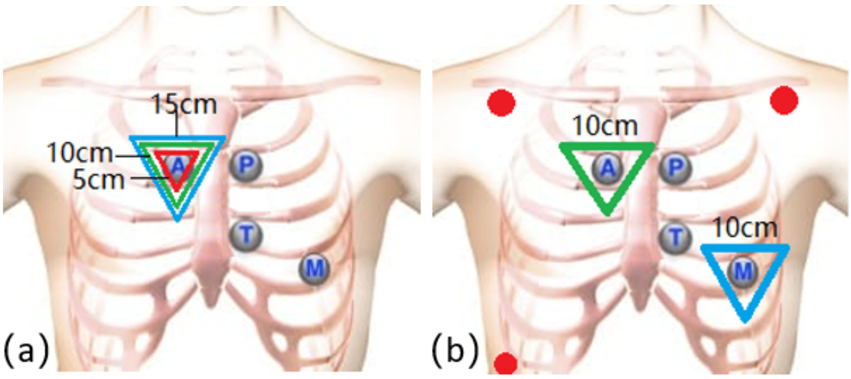


Fig. 4.2 (a) Placement of electrodes for different inter-electrode distanc (IED).
(b) Placement of sensors for different auscultation locations.

4.2.2 Signal Processing

In this study, the delays were analysed using the temporal locations of the R-peaks, the Q points and the T wave ending points in ECG, and the S1, S2 starting points in PCG. The processing was conducted in the Matlab® R2018b environment.

The captured ECG and PCG signals were filtered first to remove the unwanted noise. For the ECG, a 3^{rd}-order infinite impulse response (IIR) high-pass filter with 1 Hz is used to eliminate the baseline wander [178]. For the PCG, a 150 Hz low-pass IIR Chebyshev type I filter of order 3 is used to filter the lung sound. All filters were zero-phased.

The R-peaks, Q points and the T wave ending points are extracted from ECG signals, and S1, S2 starting points are extracted from PCG signals. Figure 4.3 shows the extraction results of the parameters.

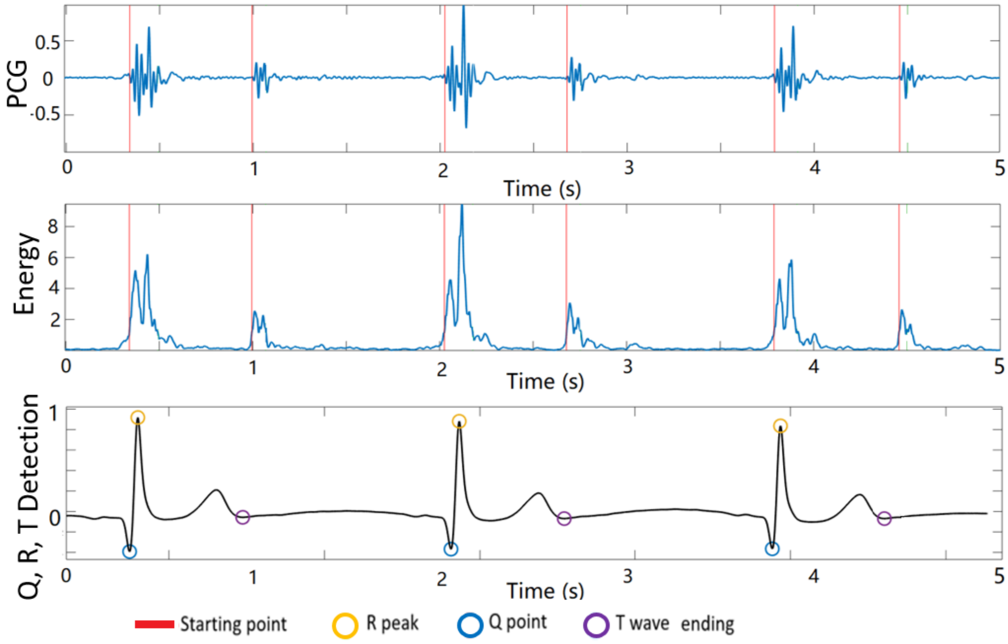


Fig. 4.3 The S1,S2 starting points in PCG, and Q, R, T points in ECG extraction result.

To capture more accurate R-peaks, the Pan-Tompkins algorithm [89] is used. The ECG signal is differentiated filtered and squared to enhance the dominant peaks (QRSs) and reduce the possibility of erroneously recognizing a T-wave as an R-peak. After the square, the R peaks of the ECG can be easily detected by setting an appropriate threshold. By using the intermediate coordinates between the R peaks, the ECG and corresponding PCG can be cut into one cardiac cycle. The data is analyzed using 5 consecutive cycles. The Q points are detected by calculating the slope on the left side of R. When the slope (first derivative) is becoming greater than or equal to 0, the first lowest point is reached, which is the Q point. The detection of T wave ending point is based on the relationship between R peak and T wave, where T wave normally occurs 250 ms – 350 ms after R peak [179]. Therefore, the peak point in this period is the T peak, and then we use the same method of Q detection to find T ending point by the first derivative.

For S1, S2 starting point detection, the short-term energy (STE) method is used. Its equation is

$$E_n = \sum_{m=0}^{N-1} x_n^2(m), \tag{4.1}$$

where E_n is the short-term energy of the signal X_n at frame n, and N is the length of the frame. In our study, the frame length is 10 ms, and frame increase is 0.5 ms. There are two thresholds to determine whether the sound is a heart sound or noise: energy threshold and duration threshold. If the STE is larger than the lower energy threshold (10% of the maximum energy), it is regarded as the potential start point. When it becomes larger than the higher energy threshold (25% of the maximum energy) and its duration is longer than the threshold, this sound is regarded as heart sound.

After the parameter extraction, the captured data are shown in Table 4.1. It is worth mentioning that during the parameter extraction, manual check is also used to reduce the error and enhance its accuracy.

Table 4.1 Extracted data name from ECG and PCG. The subscripts (Ref, A, P, T, M) mean the data is collected by the placement of the electrodes in standard Lead I or around auscultation sites.

ECG	R peak	Q point
ECG_{Ref}	R_{Ref}	Q_{Ref}
ECG_A	R_A	Q_A
ECG_P	R_P	Q_P
ECG_T	R_T	Q_T
ECG_M	R_M	Q_M
T ending point	**S1 starting point**	**S2 starting point**
T_{Ref}	–	–
T_A	$S1_A$	$S2_A$
T_P	$S1_P$	$S2_P$
T_T	$S1_T$	$S2_T$
T_M	$S1_M$	$S2_M$

4.2.3 ECG Delay Estimation Method

In this study, the Cross-correlation (CC) method is used for ECG delay estimation. CC is a function to measure the similarity of two signals by calculating the sliding inner-product, which is given as:

$$(s_1 * s_2)[\tau] \triangleq \sum_{m=-\infty}^{\infty} \overline{s_1[t]}\, s_2[t+\tau], \tag{4.2}$$

where s_1 and s_2 are the two signals to be compared, $\overline{s_1[t]}$ is the complex conjugate of $s_1[t]$, and τ is the displacement for inner-product. When $(s_1 * s_2)$ is the largest, it means the similarity is the greatest. For the ECG signals, they are regular and periodic, so the displacement to get the maximum CC is equivalent to the delay between the two signals. Using this relation, the time delay between the two ECG signals can be determined by:

$$\tau_{delay} = \arg\max_{t \in \mathbb{R}}((s_1 * s_2)(t)), \tag{4.3}$$

4.2.4 PCG Delay Calculation Method

The PCG signals are relatively complicated and not regular as ECG. Thus using the CC method causes a significant estimation error. Therefore, the delays are calculated directly by the difference of the key points. The error is reduced by calculating the mean of the five heart cycles. According to the extracted data in Table 4.1, the calculated delays are shown in Figure 4.4.

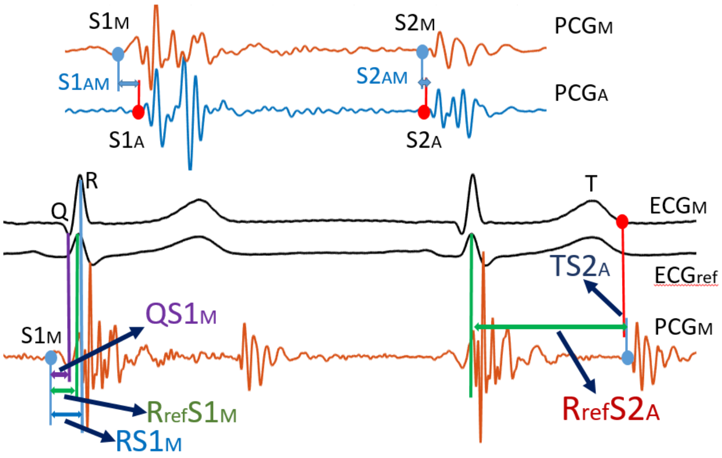


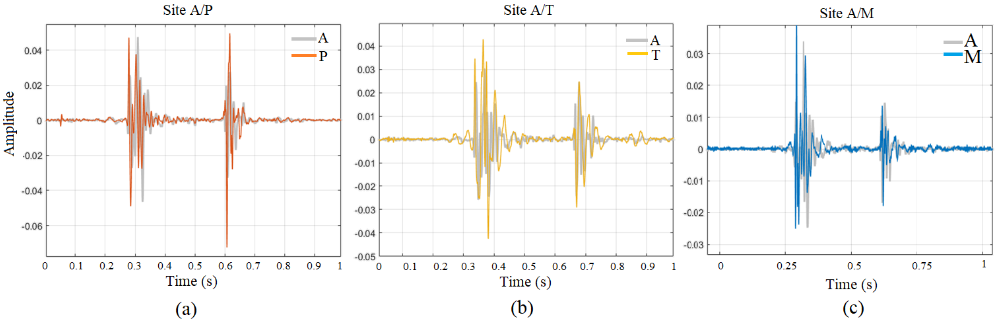
Fig. 4.4 The calculated delays associated with PCG.

$S1_{AM(AP,\ AT)}$ are the delays between S1 onset in site A and the other auscultation sites. $S2_{AM(AP,\ AT)}$ are the delays between S2 onset in site A and the other auscultation sites. $RS1_{M(A,\ P,\ T)}$ are the delays between S1 onset and R peak in each auscultation site. $R_{ref}S1_{M(A,\ P,\ T)}$ are the delays between S1 onset and R peak in reference ECG. $QS1_{M(A,\ P,\ T)}$ are the delays between S1 onset and Q point in each auscultation site. $R_{ref}S2_{A(P,\ T,\ M)}$ are the delays between S2 onset and R peak in reference ECG. $TS2_{A(P\ ,T\ ,\ M)}$ are the delays between S2 onset and T wave ending in each auscultation site.

4.3 Results

4.3.1 PCG Morphological Variation among the Auscultation Sites

Figure 4.5 shows a representative comparison of PCG signals at different auscultation sites from one healthy subject. Because the PCG were captured in pair during the experiment, they are demonstrated with site A as reference. The PCG waveform is affected by various factors, including the location of the recording site. Changing the recording site can lead to visible changes in the morphology of the PCG waveform. Theoretically, when the stethoscope is placed over the apex of the heart, which is the lower tip of the heart, the S1 heart sound is usually more prominent in the PCG waveform. The S1 heart sound is caused by the closure of the mitral and tricuspid valves at the beginning of systole (when the heart is contracting). On the other hand, when the stethoscope is placed over the base of the heart, which is the upper part of the heart, the S2 heart sound is usually more prominent in the PCG waveform. The S2 heart sound is caused by the closure of the aortic and pulmonary valves at the end of systole. As in Figure 4.5(a), the S2 amplitudes are similar as S1 amplitudes at site A and P, while in Figure 4.5(b) and (c), S1 is viablely larger than S2. There are also morphological differences in amplitude or duration. The reason of these differences is that S1 and S2 are generated by the closure of different valves. Thus, captured at specific sites may enhance the volume and affect the sound propagation.



(a) (b) (c)

Fig. 4.5 (a) PCG at site A and P. (b) PCG at A and T. (c) PCG at A and M.

4.3.2 The Effect of IED on ECG Delays

Figure 4.6 shows the IED of 5 cm and 15 cm compared with 10 cm. The delays of the occurrence of R peak are all close to 0 ms (mean ± standard deviation; 5cm: -0.359 ± 2.181 ms, 15cm: 0.805 ± 1.861 ms), except one outlier. Therefore, we can basically conclude that the IED does not affect the R peak occurrence significantly, and there is no obvious regularity in the effect.

4.3.3 Delay between Standard Lead 1 ECG and Site Specific ECG

As shown in Figure 4.7, the delay between auscultation sites shows an increasing trend from site A to site M. Compared with the standard lead I ECG, the ECG at A site is normally negative, which means the R peaks at site A is advanced, R peaks at P and T are close to standard lead I ECG; and R peak at M comes later than standard lead I ECG.

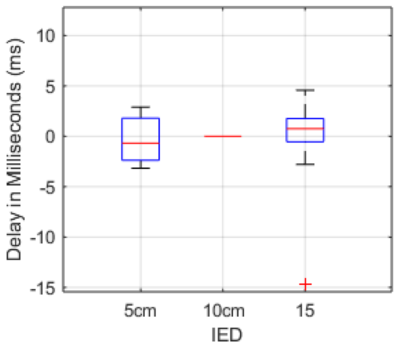


Fig. 4.6 Inter-electrode distance (IED) caused ECG delay (10 cm as reference).

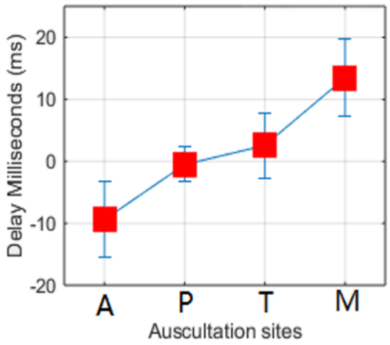


Fig. 4.7 The delay (mean ± standard deviation, SD) between auscultation sites ECG and standard Lead I.

4.3.4 PCG Delay between Site A and the Rest

For the PCG delay between the auscultation sites (site A is as the reference), S1 and S2 are analysed separately. The results are shown in Figure 4.8. S1 onset becomes earlier (negative) from A to M. However, S2 onset almost remains the same from A to T, but there is a slight delay (positive) at M.

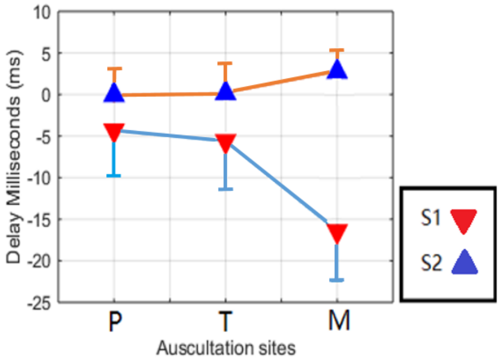


Fig. 4.8 The delay of S1 (red) and S2 (blue) in each auscultation site (A site as reference).

4.3.5 Delay between ECG and PCG

Figure 4.9 shows the delay between S1 onset in each auscultation site and R peak in standard Lead I ECG. The delay trend is similar to the S1 onset delay trend, but it can be seen that at site A, the onset of S1 occurs after R-peak. When it comes to site M, the S1 onset is basically before R-peak. Figure 4.10 illustrates the delay of S1 onset, R-peak and Q point in the auscultation site ECG. The delay trend is similar to the delay of Standard Lead I, but it becomes larger. At site A and M, there are average 20 ms time difference between S1 onset and R-peak. However, compared with Q point, the average delay in site M is close to 0.

S2 is widely regarded as occurring right after T-wave. In this study, it is found that the S2 onsets are basically after T wave ending points in auscultation area, except 4 groups of outlier as shown in Figure 4.11(a). Besides, the relationship between Lead I ECG R-peak and S2 onset is also presented in Figure 4.11(b).

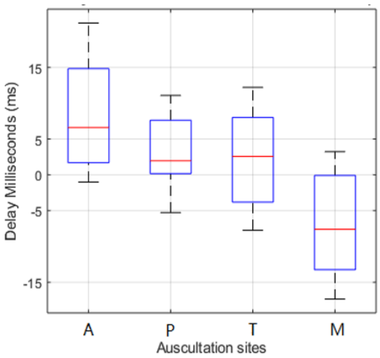


Fig. 4.9 The delay between S1 onset for each auscultation site and R peak in standard Lead I.

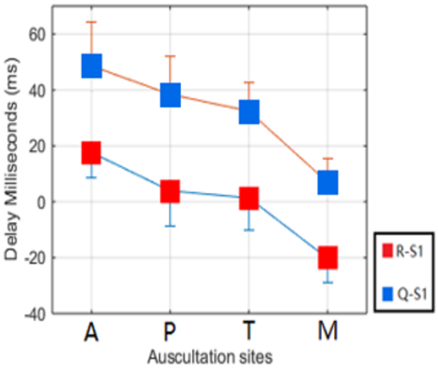


Fig. 4.10 The delay between S1 onset and R-peak/ Q-point in each auscultation site.

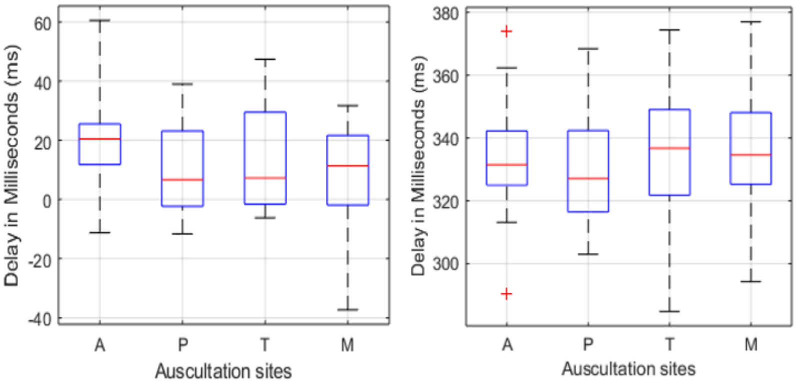


Fig. 4.11 (a) The delay between S2 onset and T wave ending point in each auscultation site. (b) The delay between S2 onset and R peak in Standard Lead I.

4.4 Discussions

This study aimed to analyse the time delay between ECG and PCG at different auscultation sites (A, P, T, M), and investigated the changes in the time occurrence of the R-peak in relation to the distance between the recording electrodes. The results shown firstly that the time property for PCG segmentation based on ECG in the previous study can be misleading and holds only for specific auscultation sites. In our recorded PCG signals, the S1 onset was gradually advanced from auscultation site A to site M, while the S2 was delayed in the meantime. This result is basically in line with our previous knowledge that S1 is generated at heart apex (site M and T) and S2 is generated at heart base (site A and P) [180], so S1 should be captured at site M earlier and S2 should be captured at site A earlier. As the result of heart sound propagation variation on the chest, the S1 onsets in the captured PCG can occur before or after R peak in the ECG. Normally, the S1 onset is after R peak at site A, before R peak at site M, and adjacent to R peak at site P and T. Therefore, distinguishing the auscultation location is necessary for doing more precise segmentation.

Secondly, there is regularity in the translation of the R-peak relative to the auscultation sites. Considering the body as a volume conductor, we can conclude the R-peak of the ECG signals is conducted from site A to site M. Because site T and M coincide with the chest lead in clinic, we also analyzed the open-source 12 leads ECG, and found analogous delay trend between V2 to V6. This finding is similar to the electrical axis caused QRS complex deviation, and the direction is reverse for PCG and thus care should be taken when using the ECG as a reference signal to segment the PCG. When the signal is captured at site M where S1 onset is far before R peak, Q point can be an alternative reference point for the segmentation.

Thirdly, the IED has not effect on the R peak shifting. Thus, shortening the IED can be of help to reduce the size when designing ECG-PCG integrated small device.

Besides, it is found that the RSR’ (An ECG finding in which there are two R waves) happened in 5 subjects’ site A ECG during the experiment. Normally the RSR’ occurs in the conditions of right bundle branch block (RBBB) or left bundle branch block (LBBB) [181], but there is no such physiological or heart conditions on the subjects. Therefore, it worth noticing to choose appropriate R peak when using ECG to do PCG segmentation under this condition. In our analysis, the first peak was used in the delay calculation and it conforms to the rest trend.

Lastly, there are also some limitations in this study. The IED effects on ECG were tested by only 5 cm, 10 cm and 15cm which was limited by the diameter of the electrodes (4 cm). If there are more interpolations between them, the result can be more convincing and accurate. In the analysis of IED effect on ECG, there is one outlier with around 15 ms R peak shifting cannot be explained. It is conjectured that the error was caused by the misplacement of the electrodes.

The study found that when the ECG is captured at auscultation sites, the R peak of ECG shifted backward regularly from A to M, and the distance between the electrodes did not affect the R peak shifting. In addition, the propagation of the heart sound on the chest caused a delay on S1 onset in the captured PCG signals. Therefore, the R peak shifting and PCG delay lead that using R peak to directly locate S1 in PCG no longer accurate. This can be improved by distinguishing the time property of each auscultation site. All the findings are of help in designing small ECG-PCG integrated device, and providing theoretical basis for using ECG to do more accurate PCG segmentation.

Chapter 5

Classification of PCG Signal Using Deep Learning Techniques

5.1 Introduction

Auscultation is the most common and practical approach in early cardiac screening. However, it is not always reliable due to physician's clinical experience and listening ability. Computer-aided heart sound analysis is a potential method to improve auscultation accuracy. It can overcome the limitation by using machine learning (ML) algorithms to learn the diagnostic ability on large amounts of data.

The first published study on automatic heart sound classification can be traced back to 1963 when research used threshold to identify rheumatic heart disease [182]. Afterwards, more ML techniques have been explored, such as logistic regression [57], regression tree [49], K-nearest neighbors (KNN) [57, 52, 59], random forest [58], support vector machine (SVM) [57, 60, 61], hidden Markov model (HMM) [63]. Besides traditional ML approaches, neural network (NN) and its variants have been also applied to heart sound classification [28, 183, 54]. With the development of NN and advanced computing hardware, more dedicated and deeper NNs have been proposed, such as convolutional neural networks (CNNs). CNN focuses on image recognition; and recurrent neural network (RNN) is specialized in prediction sequence data such as audio and text [184, 185]. Compared with traditional ML methods, deep learning algorithms can neglect manual feature extraction and use the raw signal as input with promising performance. After the 2016 PhysioNet/Computing in Cardiology (CinC) Challenge [42], using CNN or RNN to conduct heart sound classification became the mainstream approach [50, 65, 66, 68–70, 51, 186, 8, 187, 188].

This chapter introduces the three conducted studies on optimization of deep learning (DL) inputs and architectures to improve the PCG classification. Section 5.2 is the preliminary knowledge for understanding these studies. The studies includes: (1) exploring the signal duration effect on RNN and CNN performance in Section 5.3; (2) comparing the performance of different time-frequency distributions as CNN input in Section 5.4; (3) proposing novel algorithm for PhysioNet Challenges 2022 in Section 5.5. At last, Section 5.6 summarizes the three published works.

5.2 Preliminary

This section summarizes the preliminary knowledge for understanding the conducted studies, including the principal and conversion of Time-frequency distributions (TFDs), extraction of Mel-Frequency Cepstral Coefficients (MFCCs), basic concept of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) with their mainstream architectures. Besides, the PhysioNet Database for PCG classification and the evaluation metrics are also introduced to avoid repeat instructions in subsequent studies.

5.2.1 Time-Frequency Representations

Time-frequency distributions (TFDs) represent signals in the time and frequency domains, used in a wide range of signal analyses, such as optics, acoustics, and biomedicine [48]. For bio-signals, TFDs are used to analyse electroencephalography (EEG), electromyography (EMG), ECG, etc [189]. TFDs transform a 1-D signal in the time domain into a 2-D space, containing both time and frequency information. This transformation can provide a more intuitive visualisation of the frequency information. Furthermore, as inputs for CNNs, they can also reduce the difficulty of frequency information mining. For PCG classification, the murmurs are defined in time and frequency, therefore, the availability of this information to the classifier is a reasonable basis for improvement. The short-time Fourier transform (STFT) is the most commonly used transform, which is relatively simple to calculate. However, it cannot effectively track the abrupt changes because a trade-off is needed between the time and frequency windows' size. This means that an increased time-domain resolution worsens the resolution in the frequency domain and vice versa [190, 191]. The continuous wavelet transform (CWT) is another approach for time-frequency analysis. Compared with STFT, it can capture more local information by scaling and shifting the mother wavelet to fit the raw signal. This improves the resolution but could not provide greater frequency resolution at higher frequency bands [192]. However, due to the low-frequency characteristics of the heart sound, this drawback of CWT can be neglected. Thus, CWT is often used in PCG analysis [188, 193–195]. The Chirplet transform (CT) can be regarded as a generalisation of both STFT and CWT, including four parameters: time, frequency, scale, and chirp rate [191, 196, 197]. It also holds the advantage of detecting instantaneous frequency trajectory used in PCG analysis [198, 196]. In addition to the mentioned TFDs, Cohen's class distributions such as Wigner-Ville distribution (WVD) and Choi-Williams distribution (CWD) were also applied in the heart sound analysis [199, 200]. They can provide higher resolution, but with cross-term issues (also known as interference terms). In [201], several TFDs were graphically compared and analysed on their ability to visualise PCG events.

1) Short-time Fourier transform

Fourier transform is a linear integral transform that converts the signal $x(t)$ from the time domain into the frequency domain $\widehat{X}$. To provide simultaneous time and frequency information, STFT is applied with a sliding window function $\omega(t)$ on the signal conducting a fast Fourier transform (FFT) within the windows to determine the frequency variation over time. Eq. (5.1) defines the STFT:

$$\widehat{X}_{STFT} = \int_{-\infty}^{\infty} w(t-\tau)x(\tau)e^{-j\omega\tau}d\tau \tag{5.1}$$

where the transform output $\widehat{X}_{STFT}$ can be regarded as a time-dependent frequency spectrum (spectrogram), t and ω are the time and frequency of the window function. The prerequisites of using STFT are signal stationarity within the window because burst or transient signals will significantly affect the performance.

2) Continuous wavelet transform

Continuous wavelet transform (CWT) is another commonly used TFD that performs well in non-stationary signals such as EEG and ECG. It can extract better local information in both time and frequency domains. The CWT (scalogram) of signal $x(t)$ can be calculated as given in Eq. (5.2).

$$\begin{aligned} \widehat{X}_{CWT} &= \frac{1}{\sqrt{\alpha}} \int_{-\infty}^{+\infty} x(t)\psi^{*}\left(\frac{t-\beta}{\alpha}\right)dt \\ &= \sqrt{\alpha}\int_{-\infty}^{+\infty} X(\omega)\psi^{*}(\alpha\omega)e^{j\omega\beta}d\omega \end{aligned} \tag{5.2}$$

where α is the scale parameter that is inversely related to the frequency, β is the shifting parameter, $\psi^{*}(t)$ is the complex conjugate of the mother wavelet and $\psi(\frac{t-\beta}{\alpha})$ means the contracted and stretched mother wavelet for fitting the signal $x(t)$. In essence, the output of CWT is the convolution of the input signal with the mother wavelet. Thus, on the right-hand side equation, $X(t)$ and $\psi(\omega)$ are the Fourier transformed $x(t)$ and $\psi(t)$, respectively. Unlike STFT, which uses the same sliding window for all frequency bands, CWT can vary the window size by the scaling parameter. The mother wavelet is contracted when the scaling parameter is small, providing finer resolution. In contrast, the mother wavelet is stretched when the scaling parameter is large, providing coarser resolution. Because the scaling parameter can be regarded as the inverse frequency, the pseudo-frequency of the CWT can be approximated by Eq. (5.3):

$$f = \frac{f_{mv} f_s}{\alpha} \tag{5.3}$$

where f_{mv} is the centre frequency of the mother wavelet, f_s is the sampling frequency, and α stands for the scale. [188] showed that Morlet is the most appropriate mother wavelet for heart sound analysis.

3) Chirplet transform

Chirplet transform (CT) can be regarded as an improved wavelet transform by further modifying the mother wavelet. It rotates the 'wavelet' in the time-frequency plane, which is equivalent to applying a nonnegative, symmetric, and normalized window $\omega(\sigma)$ (usually a Gaussian window function) [202]. The output 'chirplet' is windowed 'wavelet' in scaling and time-shifting. In addition, another two parameters are introduced, chirping (frequency rotation operator) $\Phi_\alpha^R(t)$ and frequency shifting $\Phi_\alpha^M(t,t_0)$, calculated by Eq. (5.4):

$$\begin{aligned} \Phi_\alpha^R(t) &= e^{-j\alpha t^2/2} \\ \Phi_\alpha^M(t,t_0) &= e^{j\alpha t_0 t} \end{aligned} \tag{5.4}$$

where $j = \sqrt{-1}$, α is chirp rate, and t is time. The rotatory angle of the analytical associate of the signal is $\theta = \tan^{-1}(-\alpha)$ and the shift in the frequency is from ω to $\omega + \alpha t_0$. The analytical associate $z(t)$ of the signal can be obtained by Hilbert transform. The chirplet transform of a signal $x(t)$ can be expressed as Eq. (5.5):

$$\hat{X}_{CT}(t_0, \omega, \alpha) = \int_{-\infty}^{+\infty} \bar{z}(t) W_{(\sigma)}(t - t_0) e^{-j\omega t} dt \tag{5.5}$$

where ω is the angular frequency, $\bar{z}(t) = z(t)\Phi_\alpha^R(t)\Phi_\alpha^M(t,t_0)$. The terminology of the rest variables is the same as above.

4) Wigner-Ville distribution

Cohen's class distributions are also known as bilinear or quadratic time-frequency distributions. Wigner–Ville distribution (WVD) is one of Cohen's class's most used distributions. As described, STFT needs to balance the time and frequency resolution; however, the bilinear time-frequency distributions, such as WVD, can offer excellent resolutions in the time and frequency domain without trade-offs. They convert the time and frequency of the signal into the complex conjugate signal to present the energy distribution, and their resolutions are only related to the signal length. The calculation of Cohen's class distribution is given by Eq. (5.6):

$$C_x(t,f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A_x(\eta,\tau)\Phi(\eta,\tau)\exp(j2\pi(\eta t - \tau f))d\eta d\tau \tag{5.6}$$

where $A_x(\eta,\tau) = \int_{-\infty}^{\infty} x(t+\tau/2)x^*(t-\tau/2)e^{-j2\pi t_j}dt$ is the ambiguity function and $\Phi(\eta,\tau)$ is the kernel function (normally low-pass function) to reduce the noise. When $\Phi(\eta,\tau) = 1$ it is the WVD, expressed as in Eq. (5.7):

$$\widehat{X}_{WVD}(t,f) = \int_{-\infty}^{\infty} x(t+\tau/2)x^*(t-\tau/2)e^{-j2\pi f\tau}d\tau \tag{5.7}$$

Although WVD provides a better resolution of the signal TFD, it has the issue of cross-term (false indication of the existence of the signal components), which is a common problem of Cohen's class distributions. Therefore, the smoothed pseudo-WVD is normally selected to reduce the cross-term issue.

5) Choi-Williams distribution

The formal approach for reducing and eliminating the cross-term interference of Cohen's class distributions is to select the appropriate kernel function. However, this may also corrupt the valuable information of the signal. Besides the pseudo-WVD, the Choi-Williams distribution (CWD) with its kernel function $\Phi(\eta,\tau) = e^{-\partial(\eta\tau)^2}$ was showed its potential for PCG visualization in [200].

5.2.2 Mel-Frequency Cepstral Coefficients

Mel-frequency cepstrum (MFC) is a vital representation in sound processing that uses Mel-scale instead of linear scale to display the short-term power spectrum. The advantage of the Mel-scale is that it can reflect more closely to the human non-linear auditory system. The human auditory sense is more sensitive to identifying the voice changes in low frequency. In contrast, people need a larger frequency band in high frequency (over 500 Hz) to distinguish the same pitch increment. The relationship between the two scales (Hz and Mel) is:

$$Mel(f) = 2595log_{10}(1+\frac{f}{200}) \tag{5.8}$$

where $Mel(f)$ is the frequency in the Mel-scale, f is the frequency in the linear scale. *MFCCs* are the coefficients derived at Mel filter banks that describe the energy distributed in the different critical frequency bands. The extraction of *MFCCs* includes the following steps:

1) Pre-emphasis

The purpose of pre-emphasis is to amplify and compensate the high-frequency component that is suppressed in sound production. This will increase the Signal-to-Noise

Ratio (SNR) and balance the frequency by enhancing the high-frequency content, which is usually small in magnitudes. A pre-emphasis filter shown in Eq. (5.9) is generally applied as:

$$x(t) = s(t) - \mu s(t-1) \tag{5.9}$$

Here $s(t)$ is the sound signal, $x(t)$ is the filtered signal, and $\mu \in [0.9, 1]$ is used for high-pass pre-emphasis.

2) Fourier transform

To obtain the spectrum of the input signal, a STFT is used to convert the signal from the time domain into the frequency domain. The conversion is:

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-2j\pi nk/N}, 0 \leq k \leq N \tag{5.10}$$

where $x(n)$ is the input signal, $X(k)$ is the corresponding Fourier coefficients, and N represents the number of samples in each frame. The power spectrum is $X(k)^2$.

3) Mel-scaled power spectrum

Mel spectrum could be obtained by the power spectrum passing through a set of Mel-scaled filter banks, where the banks $B_m(k)$ are:

$$B_m(k) = \begin{cases} 0, & k < f(m-1) \\ \frac{k-f(m-1)}{f(m)-f(m-1)}, & f(m-1) \leq k \leq f(m) \\ \frac{f(m+1)-k}{f(m+1)-f(m)}, & f(m) \leq k \leq f(m+1) \\ 0, & k > f(m+1) \end{cases} \tag{5.11}$$

m is each Mel filter of all M filters. k is the samples in frames. $f(m-1)$, $f(m)$ and $f(m+1)$ represent the beginning, medium, and end frequency of each Mel triangle filter. The Mel-scaled power spectrum is the product of the power spectrum $X(k)^2$ and the banks $B_m(k)$, which is:

$$P(n) = \sum_{n=0}^{N-1} X(k)^2 e B_m(k), 0 \leq m \leq M \tag{5.12}$$

4) Discrete cosine transform

MFCCs could be derived by the DCT of the logarithmic Mel spectrum:

$$\begin{gathered} MFCCs(i) = \sum_{m=1}^{M} \log[P(n)] \cos\left[i\left(m - \frac{1}{2}\right)\frac{\pi}{M}\right], \\ i = 1, 2, 3, \ldots, L \end{gathered} \tag{5.13}$$

where L shows the order of frame for the $MFCCs$ and M represents the number of Mel filter banks.

5)Dynamic characteristic ($\Delta\&\Delta^2 MFCCs$)

Because $MFCCs$ can only describe the sound signal's static information (spectral envelope shape), heart sound is inherently time-variant, and the dynamic information may help describe the signal more accurately. In addition, the human ear is more sensitive to sound changes. Thus, to better emulate the auscultation procedure, adding the dynamic information of the heart sound signal may improve the detection accuracy. The extraction of $\Delta MFCCs$ is:

$$d_i = \frac{\sum_{n=1}^{N} n\left(MFCCs_{n+i} - MFCCs_{n-i}\right)}{2\sum_{n=1}^{N} n^2} \tag{5.14}$$

where d_i is delta coefficient between $MFCCs_{n+i}$ and $MFCCs_{n-i}$ at frame i, and N is usually set to 2. Similarly, $\Delta^2 MFCCs(Di)$ could be calculated by:

$$D_i = \frac{\sum_{n=1}^{N} n\left(d_{n+i} - d_{n-i}\right)}{2\sum_{n=1}^{N} n^2} \tag{5.15}$$

5.2.3 Convolutional Neural Networks

CNN is a deep learning model that consists of convolutional (Conv), pooling, and fully connected layers. The Conv layer has a set of spatially small and learnable filters (kernels) working as feature detectors. They move across the input matrix "like sliding window", calculating the dot product of the kernel parameters and windowed inputs. The output can be interpreted as the extracted feature map subject to the kernels. Usually, more than one kernel is used for the convolution. Thus, the output size is several times larger than the input. For instance, an input size $32 \times 32(1024)$ matrix after four kernels (size 3×3, stride 1, padding 0) convolution outputs $4 \times 30 \times 30(3600)$. Therefore, the pooling layer is used for down-sampling to reduce the feature maps and the amount of computation and control overfitting. Max pooling is the most commonly used pooling layer, which remains the maximum in the rectangular filtered region. Following the last instance, a 2×2 size max-pooling filter with a stride of two reserves only 25% of its original size, so only 800 parameters are transferred to the next layer. The fully connected layers can map the extracted features into catalogues for classification. Sometimes, a SoftMax layer follows the fully connected layer to the logits into a class probability distribution before the final output.

CNNs have developed rapidly since AlexNet surpassed the traditional methods by a large margin in ILSVRC2012, a visual recognition competition [203]. CNNs have numerous variants and differ in architecture, complexity, computation load. The mainstream architectures include MobileNetV3 [204], ResNet [205], SEResNet [206], and DenseNet [207], as described below.

1) MobileNetV3

CNNs are composed of stacked convolutional blocks generally containing convolution, normalisation, and non-linear activation layers. The depth and width of CNNs have to be large to extract useful high-level visual features for prediction. However, the expensive computational load caused by the over-expansion of the CNN scale might be unaffordable for deploying such CNNs on either personal computers or mobile devices. Small-scale CNNs can alleviate the computational load but might decrease the network performance. Hence, efficient architecture design is crucial for CNNs to be deployed in mobile devices with satisfactory performance.

MobileNetV3, proposed in 2019, is a lightweight CNN architecture that enables accurate and efficient computation in mobile devices for visual recognition tasks. Till now, it has been proved that MobileNetV3 can work efficiently in many computer vision tasks, e.g., image classification, segmentation, detection. Its efficient and accurate computation comes from some improvements, including depth-wise separable convolution, neural architecture searching (NAS) [208], NetAdapt [209], etc. In particular, depth-wise separable convolution is employed in all MobileNet series to reduce the number of trainable parameters by factorising traditional convolutions. NAS and NetAdapt are employed to optimise the network architecture at the block and layer levels. All these improvements make MobileNetV3 an efficient CNN architecture with fast interface time.

Besides, MobileNetV3 includes MobileNetV3-small and MobileNetV3-large, varying in depth and number of parameters. Hence, MobileNetV3 can act as the baseline for analysing the impact of model size in the heart sound classification.

2) ResNet

Increasing the depth can enhance CNNs' capacity, but it will bring about the vanishing gradient problem when the networks are too deep. It means the gradient becomes infinitely small via the multiple multiplications during backward propagation. Consequently, when a model goes deeper, its performance reaches saturation and often drops rapidly. This problem, also known as model degradation, occurred before ResNet was proposed in 2015.

ResNet is one of the most novel architectures in CNNs. Its essential contribution is skip-connection, which enables deep training and dramatically alleviates the degradation problem. With the operation, the function of some redundant layers turns into identity mapping when the network saturates. The gradient of a layer thus can backwards propagate directly to its previous layer without shrinking.

Although many CNNs have been used in the heart sound classification field, there is still no evidence of whether a deeper network can improve the classification performance or decrease due to model degradation. Hence, ResNet-18/34/50 is explored here to analyse the network depth impact while alleviating the degradation.

3) DenseNet

As aforementioned, ResNet builds a direct connection to allow gradient propagation between two connected layers. DenseNet makes more dense connections between all layers. In other words, the input of one layer in DenseNet contains the inputs of all its preceding layers. The dense connection improves the flow of both features and gradients, making training DNNs easier. Furthermore, deeper layers in a network can extract higher-level features, while lower-level features might also benefit the classification. The dense connection in DenseNet, combining features from all layers to form multi-scale convolutional feature maps, thus might help improve the heart sound classification. Notably, our study will conduct experiments on DenseNet-121/169 which is affordable in terms of computational cost.

4) SEResNet

Squeeze-and-Excitation (SE) is a module that improves the network representations by modelling channel-wise interdependencies of convolutional features. It squeezes the convolutional features across the spatial dimension to produce a vector describing the channel-wise global distribution. The subsequent excitation assigns modulation weights for each channel to rescale the convolutional features.

SE can easily embed to most existing CNN structures, e.g., ResNet, VGG, MobileNet, ShuffleNet, and improve their performance. However, it has not been experimented with whether SE can improve CNN performance in the heart sound classification yet. We, therefore, embed it to ResNet-18/34/50 to verify if SE can help CNNs achieve higher performance in this classification task.

5) Comparation of the CNNs Computing Load

Table 5.1 lists the number of trainable parameters and multiply-accumulates (MACs) of all implemented CNNs.

5.2.4 Recurrent Neural Networks

Unlike CNN focusing on the spatial characteristics of the input, RNNs are specialised in processing sequence data such as time series, text, and audio. Generally, RNNs conduct the same computation procedure cyclically on each segment of the sequences, and the following output is based on previous calculations. From network structure, it includes memory to store the hidden internal state h_t, which could be calculated by the previous hidden state h_{t-1} and input x_t, that

$$h_t = f_W(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \tag{5.16}$$

Table 5.1 Number of parameters (in megabyte, M) and multiply-accumulates (MACs) (in gigabyte, G) of the participated CNNs.

Model	Params (M)	MACs (G)
MobileNetV3-small	1.52	0.06
MobileNetV3-large	4.20	0.22
ResNet18	11.18	1.82
ResNet34	21.29	3.67
ResNet50	23.51	4.11
SEResNet18	11.27	1.82
SEResNet34	21.45	3.67
SEResNet50	26.04	4.11
DenseNet121	6.87	2.83
DenseNet169	12.33	3.36

where f_W refers to the hidden layer function such as a tanh activation function with parameter W shared across time (i.e., W_{xh} indicates the weight of the input-hidden layer, W_{hh} is the weight of the hidden-to-hidden layer, and W_yh is the weight of hidden-to-output). b is the corresponding bias vector. The predicted output is:

$$y_t = W_{yh}h_t + b_y \tag{5.17}$$

However, this general architecture faces exploding weights and vanishing gradient issues on long-term sequences. Therefore, methods such as LSTM were proposed to improve this condition.

1) Long Short-Term Memory

LSTM reminds the general RNN architecture and changes the memory cell unit structure, making it capable of storing extended time intervals. As shown in Figure 5.1(a), input, output, and forget gates control the information flow within the memory cell. For a given input x_t at specific time t, the corresponding output after passing the gates are:

$$\begin{aligned} I_t &= \sigma\left(W_{IF}x_t + U_{hI}h_{t-1} + W_{CI}C_t + b_I\right) \\ O_t &= \sigma\left(W_{xO}x_t + U_{hO}h_{t-1} + W_{CO}C_t + b_O\right) \\ F_t &= \sigma\left(W_{xF}x_t + U_{hF}h_{t-1} + W_{CF}C_t + b_F\right) \end{aligned} \tag{5.18}$$

where I, O, and F represent the input, output, and forget gate, respectively. Similar to Eq.(5.16) and (5.17), W is the weight of recurrent connections (i.e., W_{IF} indicates the weight of the input-forget gates layer). h is the hidden state, and b is the bias. σ is the sigmoid activation function. The memory C_t of this unit can be obtained by:

$$C_t = F_tC_{t-1} + I_t tanh(W_{xC}x_t + W_{hC}h_{t-1} + b_C) \tag{5.19}$$

The hidden state vector (output vector) of this LSTM unit will be transferred to the next time interval, and it can be calculated by:

$$h_t = O_t tanh(C_t) \tag{5.20}$$

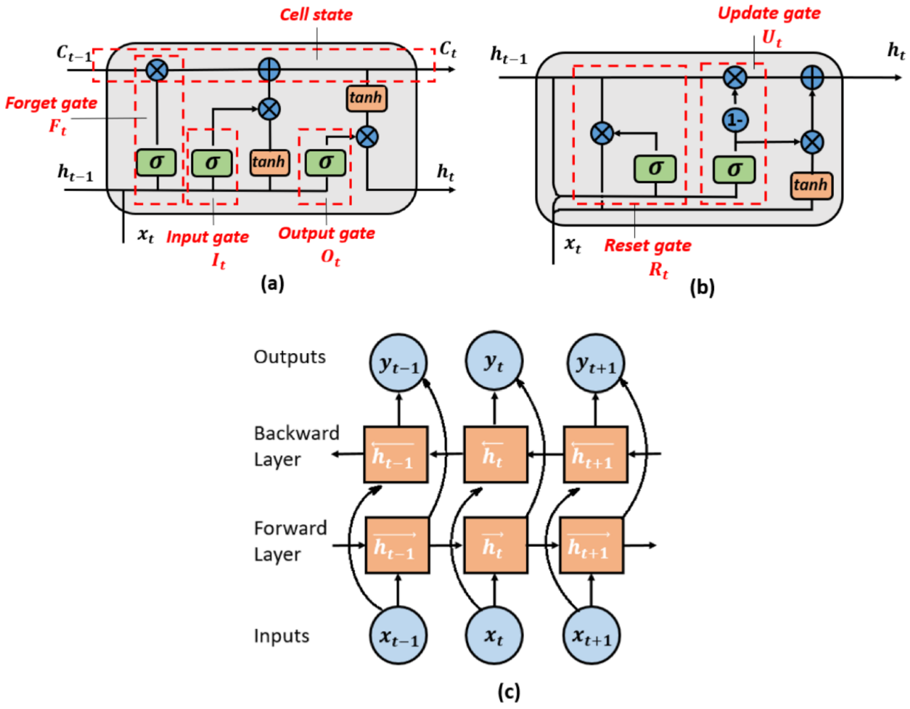


Fig. 5.1 Graphical structure of (a) Long Short-Term Memory unit; (b) Gated Recurrent Unit; (c) Bidirectional Recurrent Neural Networks.

2) Gated Recurrent Unit

GRU can be regarded as a simplified LSTM unit that integrates the input gate with the forget gate, forming a new update gate to decide the acceptance or abandonment of the information. In addition, there is a reset gate to determine how much memory is to be forgotten. The two gates work together to adaptively remember or forget during the sequence reading. Its structure is shown in Figure 5.1(b). The computational flow in one unit is given below:

$$\begin{aligned}
U_t &= \sigma\left(W_{xU}x_t + W_{hU}h_{t-1} + b_U\right)\\
R_t &= \sigma\left(W_{xR}x_t + W_{hR}h_{t-1} + b_R\right)\\
\hat{h}_t &= tanh\left(W_{\hat{h}x}x_t + U_{\hat{h}\hat{h}_t}\left(R_t h_{t-1}\right) + b_{\hat{h}}\right)\\
h_t &= \left(1 - U_t\right)\hat{h}_{t-1} + U_t\hat{h}_t
\end{aligned} \tag{5.21}$$

where U and R represent the update and reset gate. $\hat{h}$ and h are the candidate activation vector and output vector. The nomenclature of the rest variables is the same as the equations before.

3) Bidirectional Recurrent Neural Networks

Standard RNNs are unidirectional with the constraint that they can only predict the current state based on previous information. Because future information is not reachable at that moment, bidirectional RNNs were proposed to improve this situation by connecting opposite directional hidden layers to the same input. As a result, the output layer can obtain both previous and future states information by the forward and backward pass. The structure of the BiRNN is shown in Figure 5.1(c).

5.2.5 Datasets and Performance Metrics

The dataset used in Section 5.3 and Section 5.4 is from the PhysioNet 2016 database, which consists of 3153 recordings, including 2488 normal and 665 abnormal cases. They were recorded by different research teams using different electronic stethoscopes under both clinical and non-clinical settings. Because of the uncontrolled measuring environment, the duration of the recordings ranged from 5 to 120 s. Different noise types such as body motion, ambient noise, and inside body sound (i.e., intestinal sound) were added to the original heart sound. This fits the actual auscultation situation but causes more difficulty to the classification algorithm. In addition, the subjects included children, adults, and the elderly. The abnormal cases involve various heart conditions, especially coronary heart disease and valvular diseases. More details about it can be seen in [42].

Accuracy is the key metric for evaluating the performance of a classification algorithm. However, the data structure in this study was not balanced (normal: abnormal is approximately 3 : 1), so the true positive rate (sensitivity, *Se*), true negative rate (specificity, *Sp*), and overall score (*MAcc*) were also calculated as in Eq. (5.22) to obtain balanced results.

$$
\begin{aligned}
Acc &= \frac{TP+TN}{TP+FP+TN+FN} \\
Se &= \frac{TP}{TP+FN} \\
Sp &= \frac{TN}{TN+FP} \\
MAcc &= \frac{Se+Sp}{2}
\end{aligned}
\tag{5.22}
$$

where TP (True Positive) and TN (True Negative) denote the correctly classified healthy and unhealthy condition cases, respectively. Similarly, FP (False Positive) and FN (False Negative) represent the false detection on the normal and abnormal sets, respectively. The overall score (*MAcc*) is the average of the *Se* and Sp.

The dataset and metrics for George B. Moody PhysioNet Challenges 2022 are describe independently within Section 5.5.

5.3 Mel-Frequency Cepstral Coefficients of PCG Signals: Classification with Recurrent Neural Networks

5.3.1 Introduction

As described in Section 2.1, computer-aided heart sound analysis has great potential to improve auscultation accuracy by overcoming human hearing limitations and clinical experience. An electronic stethoscope is used instead of the traditional acoustic stethoscope to record the heart sound signal. Then, a machine learning classification algorithm can conduct the automated diagnosis. Typical PCG signal classification algorithms include three significant steps: segmentation (including component identification), feature extraction, and classification [210]. Firstly, the conventional segmentation breaks the whole signal into each heart cycle and locates the heart sound components. After segmentation, features such as time-frequency, energy-based, wavelet transform, and MFCCs are extracted from the signals to train a classifier. MFCCs are the most commonly used features in sound processing studies. Their frequency bands in Mel-scale can approximate human auditory system response more closely than linear-scaled spectrums because they consider the human ear perception sensitivity concerning the changing frequencies. Thus, using MFCCs as features is particularly suitable for simulating the auscultation activity. Afterwards, machine learning methods classify the input heart sounds into normal and abnormal classes based on these features.

It is still an open question on the necessity of segmentation. As mentioned, conventional segmentation aims to locate the heart sound components, helping extract the features, especially in the time domain. However, there is no widely recognised golden standard PCG segmentation technique. Unsuccessful segmentation conversely affect the accuracy and robustness of the algorithm and increase computation load. In addition, several representative studies without conventional segmentation have achieved considerable performance in the heart sound classification [68, 211, 212, 70, 50, 213, 67]. Thus, segmentation becomes an optional step that should depend on the selected features and classifiers. Generally, algorithms without segmentation make use of deep learning methods based on ANNs such as CNNs and RNNs. The signal is not segmented to identify components in these algorithms but broken into signal pieces of equal duration. As a result, different studies cut the heart sound recordings into various epoch durations, e.g., some studies broke the recordings into one-second epochs to enlarge the dataset as much as possible for training and validation [70, 67]. In contrast, others were cut into five-second pieces to retain more information in each segment [50, 65]. Different data structures and settings make it hard to evaluate the proposed methods, so it is necessary to unravel how to choose the appropriate PCG input duration for a given dataset.

The aim of this study is threefold: (1) to systematically investigate how the duration of input PCG signals will affect the performance of deep learning methods; (2) to compare the performance of commonly used RNNs, including GRU, LSTM, Bidirectional LSTM, and GRU (BiLSTM and BiGRU) under different epoch lengths; (3) to ascertain if adding dynamic information (deltas and delta-deltas of MFCCs) as the feature can improve the performance of the tested deep learning methods. Ultimately, the findings of this study will provide insight to determine reasonable input length, classifiers, and features for designing deep learning PCG classification algorithms in the future.

5.3.2 Related Work

The earliest publication on automated PCG classification can be traced to 1963, when Gerbarg et al. used a threshold-based method to identify rheumatic heart disease [182]. Since then, many articles have been published on the PCG segmentation techniques, features selection, and classification methods. The proposed classification algorithms include logistic regression [57], random forest [58], KNN [57, 52, 59], regression tree [49], SVM [57, 60, 61], HMM [63], and ANN and its variants [28, 64, 54]]. However, it was almost impossible to systematically and uniformly evaluate and compare the early research performance in this field, as they used different datasets with different classification tasks. With the development of deep learning techniques in recent years, more researchers switched from traditional machine learning to deep learning methods to design improved classification algorithms.

In 2016, PhysioNet/Computing in Cardiology (CinC) Challenge created an extensive database sourcing from nine different heart sound databases with 4430 recordings collected from 1072 healthy participants and patients with a variety of conditions [42]. This database can be used for binary classification (normal and abnormal heart sound) and as the platform to assess classifiers and features objectively. Furthermore, it provided the basis for designing deep learning PCG classification methods. Despite ambiguous discussion on the best classifier or the selection of features according to the PhysioNet challenge results, almost half of the top performances, especially the top three applied the neural network algorithms. This showed a great potential of deep learning in improving the performance of automated PCG classification. For the feature selection, 4 of the top 10 algorithms used MFCCs, given the proven track of their reliability and universality [214].

Moreover, Yang and Hsieh used RNN without segmentation, ranking 13 in the Challenge (Acc: 79%) [215]. This method aroused interest in the necessity of segmentation and led to future research on PCG classification using RNNs models. Practically, RNNs are suitable for sound recognition with the advantage of exhibiting temporal dynamic characteristics, and it has achieved great success in automatic handwriting and speech recognition [216]. After the challenge, more researchers explored deep learning methods without segmentation using the PhysioNet database. Table 5.2 outlines several representative studies in recent years.

Table 5.2 Recent advancements in heart anomaly detection using deep learning.
TF: Time-Frequency, L: Window length, S: Window shift

Authors	Year	Segmented	Features	Model	Acc (%)
Huai et al.[50]	2020	No. (L:5 s,S: 2 s)	TF	CNN + LSTM	91.06
Deng et al.[65]	2020	No. (L:5 s)	MFCCs	CNN + LSTM	98.34
Xiao et al.[66]	2020	No. (L:3 s,S: 1 s)	MFCCs, PSDs	CNN	93
Dissanayake et al.[67]	2020	No. (L:1 s,S: 0.1 s)	MFCCs	CNN, LSTM	99.72
Zhang et al.[68]	2019	No. (L:2 s)	TQP Features	LSTM	94.66
Latif et al.[69]	2018	Yes. (2,5,8 cycles)	MFCCs	RNNs	98.61
Maknickas et al.[70]	2017	No. (128 frames)	MFCCs	CNN	84.15
Rubin et al.[51]	2017	Yes. (L:3 s)	TF, MFCCs	CNN	84

From Table 5.2, MFCCs could be considered the most frequent and mainstream recognised feature. It is still controversial whether CNN or RNN (LSTM, BiLSTM, and GRU) is more suitable for heart sound classification. A combined classifier seems to be the future trend that could contain the advantages of both CNN and RNNs. Because the data length of the recordings in the PhysioNet database varies from 5 s to 2 min, different studies have cut the raw data into smaller lengths without segmentation. They have then rebuilt the pre-processed datasets for training, validation, and testing. This unifies the data, enlarges the training datasets, and improves the classification performance to a certain extent. Typically, the epoch durations ranged from 1 s (approximately one heart cycle) to 5 s (shortest heart sound recording in the PhysioNet database). However, various data structures make it hard to compare their claimed performance objectively. Moreover, it is unclear if the best input duration can ensure that both training sample size and each sample's information are sufficient.

This study fills the mentioned knowledge gap by breaking the PhysioNet datasets into different lengths ($1-5$ s) to rebuild new datasets. Inspired by the previous studies, MFCCs are chosen as features. Commonly used CNN and RNNs structures will be compared under similar settings. The study will assess the effect of input length and classifier selection on model performance.

5.3.3 Methodology

1) Pre-processing

As the principal objective of this study is to investigate the PCG duration effect on the classification performance, the raw heart sound recordings in the database are further cut into 1 s (71,344 segments), 2 s (34,982 segments), 3 s (22,510 segments), 4 s (16,749 segments), and 5 s (13,015 segments) length without overlapping. Their labels are also generated according to the raw database into normal and abnormal (two classes). Since the testing datasets in the PhysioNet Challenge are not published, we divided the available datasets into training: validation: testing by 8 : 1 : 1.

2) Feature Extraction

This study used MFCCs as the input feature for the deep learning models. The window framing for the extraction is a hamming window with 30 ms length and 20 ms overlap. Thirteen MFCCs are extracted for each window. The final feature map is a 13N × (100D - 2) matrix, where N = 1 when only MFCCs are extracted and N = 3 when deltas and delta–deltas are extracted as features. D is the duration length in second. Fig. 5.2 represents the MFCCs features from 1 s heart sound recording under both healthy and unhealthy conditions.

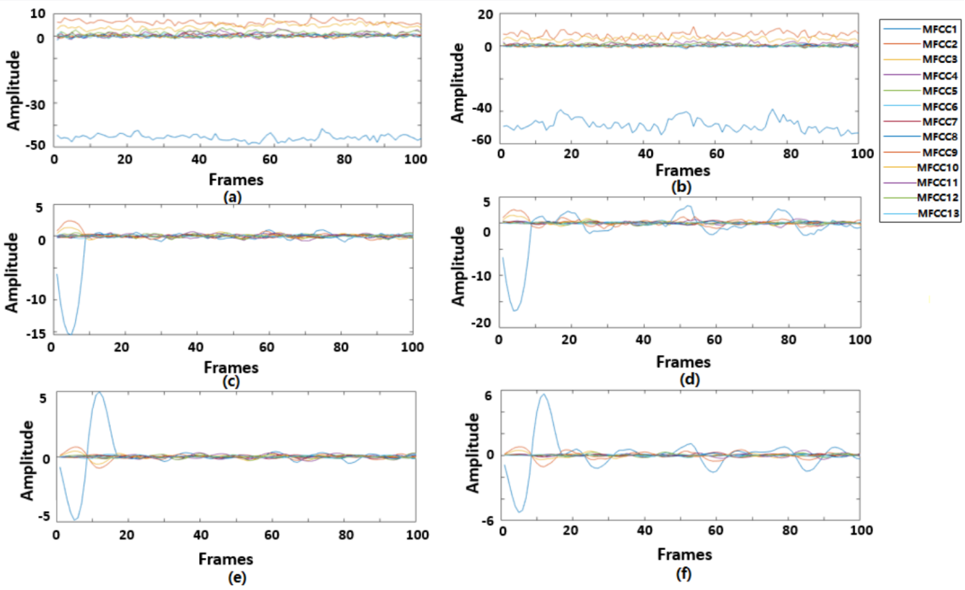


Fig. 5.2 (a) MFCCs, (c) ΔMFCCs, (e) Δ^2MFCCs for normal heart sound recording. (b) MFCCs, (d) ΔMFCCs, (f) Δ^2MFCCs for abnormal heart sound recording.

3) Model Interpretation

Another objective of this study is to compare the performance of deep learning models under similar conditions. Therefore, we built neural networks using different network structures, as shown in Fig. 5.3. The overall design is MFCCs features input to the specific network module followed by a classification module to predict.

We applied three convolutional (Conv) layers with 32, 64, and 128 filters (size 3×3) on each CNN layer. Each Conv layer is connected to a batch normalisation layer (scale 1, offset 0, momentum rate 0.9) to speed up the training and reduce the sensitivity to network initialisation before the activation function (ReLU). Two max-pooling layers (size 2) were used before the second and third Conv layers to reduce the calculation amount. After the layers, the spatial features of the input were extracted and transferred to the fully-connected layer for classification.

The RNNs models used in this study were two layers. Because in our testing, the one-layer model did not perform well on the heart sound classification with approximately 75% accuracy (two layers performed around 90%). This testing result stands in line with the description in [69]. Furthermore, deeper layers do not show noticeable improvement in the performance as well. The number of hidden units was all set to 50 in each RNN layer, the state activation function was tanh and gate activation function was sigmoid.

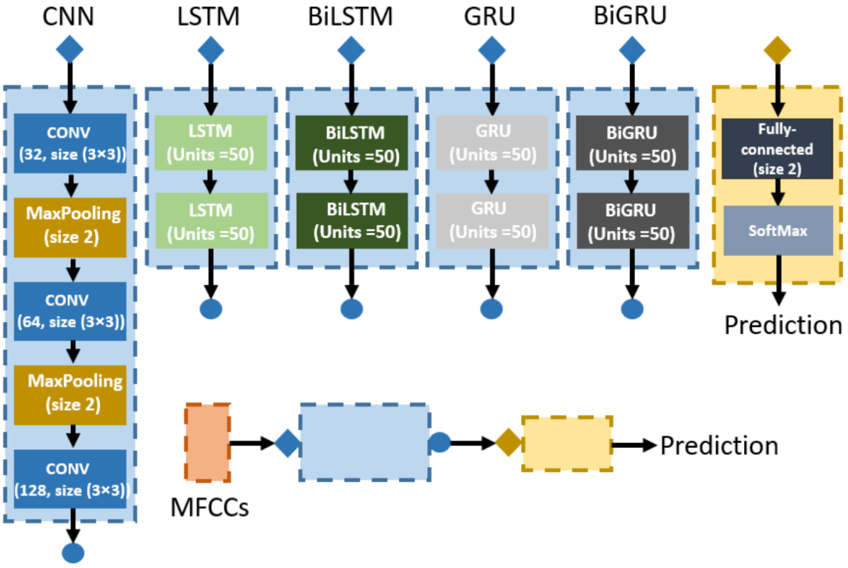


Fig. 5.3 The deep learning network models used in the study: 3-layer CNN model, 2-layer LSTM, BiLSTM, GRU, and BiGRU models.

4) Training Settings

The optimiser selected in this study is stochastic gradient descent with momentum (SGDM, learn rate 0.01, momentum 0.9). Compared with the commonly used Adam optimiser, its convergence speed may be slower, but its convergent result can be better to find the best solution. The learning rate was constant 0.01, and data shuffling was conducted for each epoch. Max epoch was set to 100, but an early stopping was applied with the patience of 5 epochs to prevent overfitting.

5) Statistical Analysis

To statistically analyse the PCG duration effect and compare the performances of the models, we trained and tested all models ten times with different input PCG lengths to avoid random results. For each time, all the models are shared with the same random seed for the division of samples into training, validation, and testing sets to guarantee comparability of the results. A non-parametric test (Mann–Whitney U test) were conducted between results for statistical purposes.

5.3.4 Results

1) PCG Duration Effect on the Deep Learning Performance

The performances (in MAcc) of the proposed models using MFCCs, ΔMFCCs, and Δ^2MFCCs as input are shown in Fig. 5.4. The RNN models (LSTM, GRU, BiLSTM, and BiGRU) showed an apparent increase between 1 and 2 s (from approximately 0.87 to 0.89), whereas there was no evident change trend between 2 and 5 s. Taking BiLSTM data as an example for statistical analysis, the p-value of 1 and 2 s is 0.017, and it is over 0.45 among 2 to 5 s, which proved that our finding is not random. However, on the CNN model, the effect of PCG duration is negligible from our results ($p > 0.3$ between different durations). Thus, we can conclude that the 1 s length PCG segment is unsuitable for training RNN models to classify heart sound, but it is acceptable for CNN models.

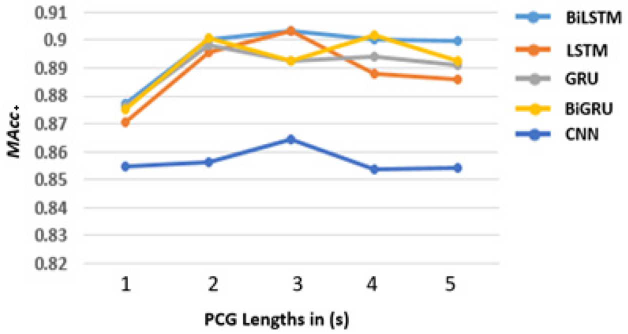


Fig. 5.4 The proposed models' performance (10 times average) with different input PCG signal lengths.

2) RNNs vs. CNN

Because there is no apparent PCG duration effect shown in Fig. 5.4 between 2 and 5 s, in this part, we analyse the performances of all the proposed models based on 5 s PCG duration, which are summarised in Table 3. RNN models outperformed the CNN model with the higher average accuracy, sensitivity, and overall score. Furthermore, the difference among the four RNN models was negligible within 1.5%($p \geq 0.175$). The best performance came from GRU with 94.07% in Acc, 94.81% in Sp, 91.29% in Se, and

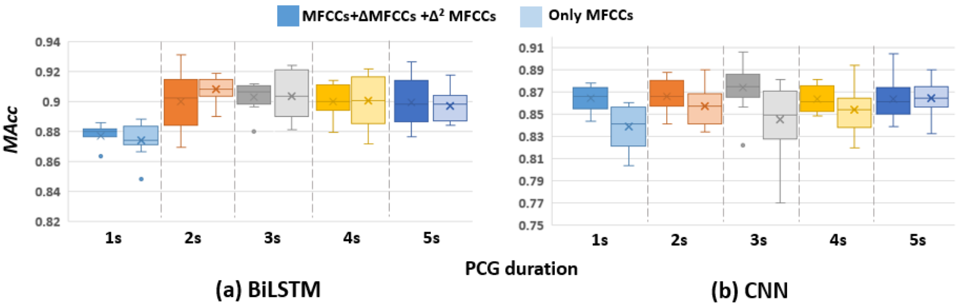
93.05% in MAcc. According to the current result, it is hard to say which RNN model is the best for heart sound classification. Notably, RNNs had better overall performance than the CNN model using MFCCs as input ($p \leq 0.044$ paired with RNN models).

Table 5.3 Comparison between the deep learning models with 5 s PCG input (10 times, average ± standard deviation %).

Model	Acc	Sp	Se	MAcc
LSTM	91.86 ± 1.20	**95.42 ± 1.45**	81.75 ± 5.95	88.58 ± 2.44
BiLSTM	**92.64 ± 0.75**	95.14 ± 1.33	**84.77 ± 3.87**	**89.95 ± 1.58**
GRU	92.13 ± 0.44	95.22 ± 1.37	83.01 ± 2.69	89.12 ± 0.87
BiGRU	92.35 ± 0.72	95.31 ± 1.87	83.24 ± 3.46	89.27 ± 1.04
CNN	90.08 ± 1.22	93.80 ± 2.46	79.02 ± 4.57	86.41 ± 1.7

3) Effect of Using ΔMFCCs and Δ^2MFCCs as Features

The performance of adding ΔMFCCs and Δ^2MFCCs as features are shown in Fig. 5.5. Because the result is applicable on all RNN models, hereby in Fig. 5.4, only BiLSTM performance is selected to display. It indicates no observed improvement on the RNN models by using extended MFCCs as the feature ($p = 0.748$). However, on the CNN model, using ΔMFCCs and Δ^2MFCCs increased the classification accuracy a little bit ($\leq 2.5\%, p = 0.003$).



(a) BiLSTM

(b) CNN

Fig. 5.5 Comparison between using mel-frequency cepstral coefficients only and using mel-frequency cepstral coefficients with its deltas and delta–deltas on bidirectional long short-term memory and convolutional neural networks.

5.3.5 Discussion

This study analyses the PCG duration effect on the heart sound classification performance and compares the deep learning models using MFCCs as input features. Because of the machine learning algorithm design, when limited training data are given, it is a standard practice to segment the database further to increase the number of training sets for accuracy improvement. Conventional segmentation (identification) breaks the PCG signals into each heart cycle for heart sound classification. However, segmentation is complicated work and cannot always complete the task accurately, especially when there is a murmur or noise inside. More studies just clipped the datasets by seconds in recent years, but no detailed analysis on the effect of the split length. In this study, the results have shown that 1 s length PCG is not an appropriate length on RNN models, while it might be applicable for the CNN model. A normal heartbeat at rest ranges from 60 to 100 per minute, which means 1 s length can only cover a complete heart cycle. For RNNs, they are specialised in processing sequence data. Thus, the whole heart cycle will be helpful to provide more comprehensive information for the trained network. However, the CNN model has the advantage of exploring specific spatial characteristics, which means missing partial information, mainly the edge information, may not directly affect the final performance of the classification. From 2 to 5 s, there is a balance between the information amount of single data and total sample size that did not affect the classification results. As a result, we suggest using 2 s PCG length to process the datasets on RNN models because a shorter segment means the potential for more repeated testing on one testing recording. This may be helpful to reduce the random error increasing the algorithm robustness, also it could control the sensitivity and specificity by appropriate threshold.

The comparison between deep learning models showed that RNNs performed better than the CNN model when using MFCCs as input. Because the MFCCs as a sequence to describe the instantaneous, spectral envelope shape of the heart sound signal did not hold too much spatial information for CNN to extract. The CNN model may perform better with a deeper structure or analyse the pattern-based time-frequency features such as the heart sound spectrogram, Mel-frequency cepstrum, etc. However, it will also require greater computing power to process during training. Among the RNN models, though their performances were quite close in our study, we suggest GRU based on the computing amount and processing time. In our testing (CPU: Intel(R) ES 1650 $v3$ @3.50 GHz, GPU: NVIDIA Quadro $M2000$), using GRU to finish the training (5 s, MFCCs only, 30 epochs) costs approximately 395 s, while LSTM needed 433 s and BiLSTM spent 588 s. From the structure, GRU is simpler than LSTM with fewer gates to calculate, bidirectional RNNs almost double the computing load by calculating the inverse propagation. Therefore, a more straightforward structure neural unit was recommended when no noticeable performance improvements were shown among them.

Thirdly, in exploring using ΔMFCCs and Δ^2MFCCs as additional features, we found no improvement on the final classification results of RNN models, but a slight increase in CNN models. This may be because ΔMFCCs and Δ^2MFCCs are derivatives of MFCCs, for RNN input sequence, no additional information was added. However, the CNN model viewed the input as a pattern; extended features can supplement more spatial characteristics. Another interesting finding is that most MFCCs information to classify heart sound as normal or abnormal was based on the first three dimensions (MFCC1, MFCC2, and MFCC3), as shown in Fig. 5.2. These dimensions correspond to the low-frequency band part of the signal. We tested using the first three MFCCs instead of all 13 MFCCs as input and found Acc 91.43%, Sp 94.39%, Se $82.07\%, andMAcc88.23\%$ on 5 s BiLSTM, which did not show an apparent decrease in the performance. Thus, it is a potential way to reduce the calculation amount.

This study found that the PCG duration will affect the deep learning performance, and the commonly used 1 s length is not a reasonable option to process the datasets. We suggested starting from 2 s since a bit longer duration can provide more information and benefit the classification performance. However, only increasing the input length without changing network architecture does not guarantee better performance. When using MFCCs as training features, RNNs outperformed the CNN model, whereas there is no apparent difference among the RNN models (LSTM, BiLSTM, GRU, or BiGRU, within 1.5%). In comparison, GRU has the advantages of a smaller computational load and a faster training speed. For MFCC features, adding dynamic information (ΔMFCCs and Δ^2MFCCs) of the heart sound did not improve the RNN performance, and the improvement on CNN is also minimal.

5.4 Time-Frequency Distributions of PCG Signals: Classification by Convolutional Neural Networks

5.4.1 Introduction

With the development of computer vision, deep CNN is partially undertaking the analysis task to provide the auxiliary diagnosis. The TFDs were used as inputs to train the deep learning algorithms [50, 217–221, 186, 8, 187]. However, due to the different databases, inputs, and network architectures, it is unclear how selecting TFDs, and CNNs can affect heart sound classification. Furthermore, many studies have proved that combining different signal processing methods can improve classification performance [222–227]. Though, their combining methods differ significantly, including channel-wise stacking [225], spatial concatenation [222], hidden feature fusing [225–227] and input vector concatenation [228]. Among the combination methods, channel-wise stacking is the mainstream.

Nevertheless, this approach has received less attention in the heart sound classification field. Hence, it is necessary to investigate whether the channel-wise stacking method to combine the different TFDs can further improve the heart sound classification.

Overall, the aims of this study are: 1) to assess if the selection of the TFDs will affect heart sound classification; 2) to compare the performance of several state-of-the-art

CNNs with different capacity, depth, and architecture, for different TFDs; 3) to ascertain if combined TFDs instead of single TFD as input can improve classification accuracy. Thus, the main contribution of this study is to provide insight into the selection of CNN inputs (in terms of TFDs) when designing heart sound classification methods.

5.4.2 Methodology

1) Pre-Processing

Referring to [229], the recordings were further segmented into 5s durations without pre-processing overlapping, generating 13015 segments (9857 normal and 3158 abnormal). Since the testing datasets were not included in the published Physionet challenge database, we randomly divided the generated segments into training, validation, and testing sets by $8:1:1$. In all three sets, the distribution of the normal and abnormal segments retained approximately $3:1$.

2) **TFDs Extraction and CNN Models Selection**

In this study, five TFD approaches were encompassed and compared as CNN inputs, including STFT, CWT, CT, WVD, and CWD. The CNN models were selected with different complexity and computing load, including MobileNetV3-Small/Large, ResNet-18/34/50, SEResNet-18/34/50, DenseNet-121/169. The theories and properties of the TFD methods have been provided in 5.2.1 and description of the CNNs models is included in 5.2.3. So here, only introduce the experimental settings.

For STFT, a Hann window of length 128 ms was applied. The overlap length was 125 ms, and the FFT was 512 samples long. For CWT, Morlet was chosen was mother wavelet according to [188]. A Gaussian window of 64 samples was applied for CT, and the number of frequency axis points associated with the spectrum was set to 512. For WVD, the smoothed pseudo-WVD was selected to reduce the cross-term issue. For CWD, $\partial = 3$ according to [200], when the TFD can clearly show the time, frequency, and intensity of the heart sound component.

Before the generated TFDs inputting into the CNNs, the TFD settings were also manually checked the clarity of visualisation. Fig. 5.6 shows the generated TFDs from both healthy and unhealthy subjects.

3) Training Settings

This study kept all the hyper-parameters consistent in all experiments for a fair comparison. The computer used one Nvidia Tesla V100 GPU card. The CNN training codes are implemented by Pytorch 1.9 [230]. All CNN backbones were implemented by the Timm library [231]. The training procedure was accelerated by automatic mixed precision (AMP) in Pytorch to save the computing memory. The results were obtained by averaging

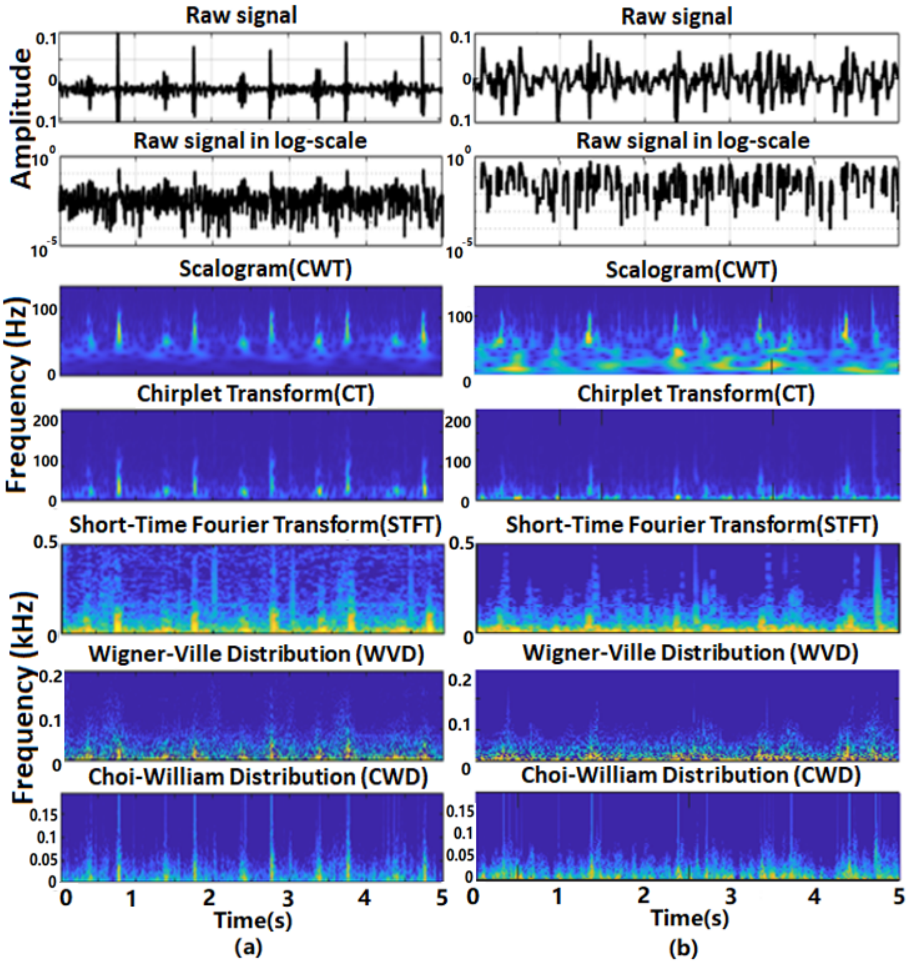


Fig. 5.6 Visualisation of the generated TFDs. (Left) Healthy heart sound. (Right) Unhealthy heart sound.

the measurements in 10 repeated experiments with different random seeds to make the comparison more convincing and alleviate randomness concerns.

The input signal images for CNNs were resized to a standard resolution of 224×224 in three channels (RGB), and the pixel values were normalised with zero mean and unit standard deviation. For raw signals and their log-scale form, their 2-D images were generated by projecting the sequential signal amplitude in y-axis, connecting neighboring points (equivalent to the “Plot” function), getting frame to obtain the waveform, and resizing the figure to standard resolution. Both single TFD and their combination were tested as CNN inputs. Specifically, we loaded three TFDs in greyscale and stacked the three greyscale inputs in channel dimension. The dimension of the combined TFDs is consistent with that of one single TFD, so no modification to the network architectures to fit the input size was needed.

The employed optimiser is Adamw with 10^{-2} weight decay. According to the cosine annealing decay schedule, the learning rate decays from an initial 10^{-2} to the minimum

10^{-6} in 50 epochs. The batch size is set to 128 to ensure batch variety and avoid memory overflow.

The performance of TFDs was compared with the baseline using original or log-scaled raw signal as CNN inputs.

4) Statistical Analysis

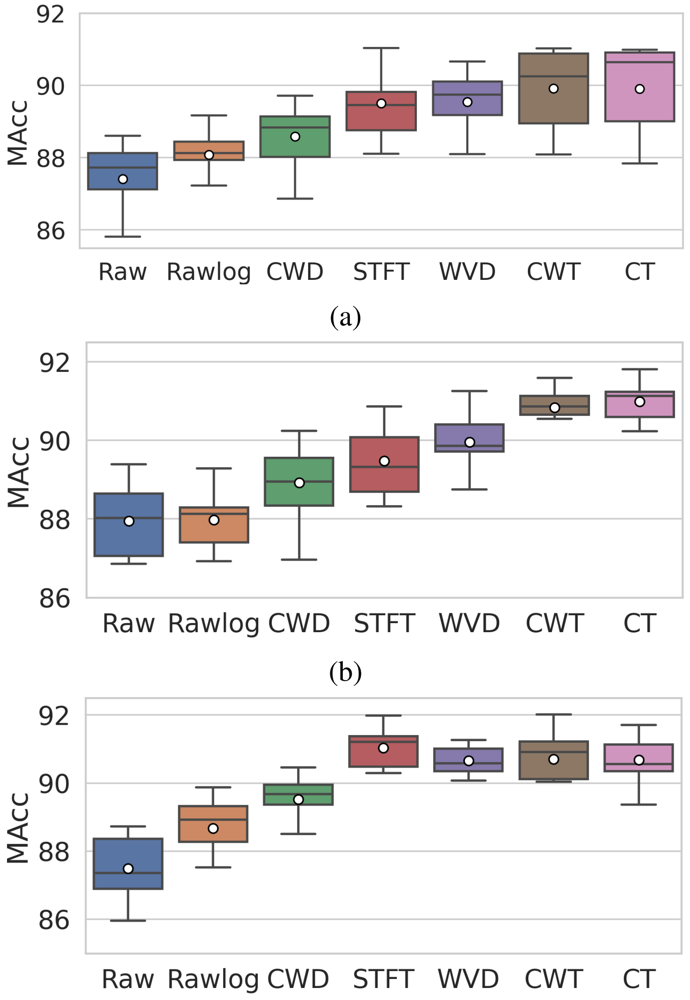
All CNNs were trained and tested ten times with different TFDs. In particular, the CNN training processes were set to be deterministic but with ten random seeds. The non-parametric test (Mann-Whitney U test) was conducted for statistical significance based on the experimental results. The p-value of the test can reveal whether a TFD or CNN outperforms the other objectively. If the p-value is more significant than 0.05, there is no significant difference between two TFDs or CNNs.

5.4.3 Results

1) TFD effects on the CNN performance

The overall performances (in MAcc, an average of 10 times) of the TFDs for 10 CNNs are illustrated in Fig. 5.7a. It revealed that transforming from the 1-D time domain to the 2-D time-frequency domain by TFDs can stably improve the classification performance. In particular, compared with the baseline (raw signal, 87.4%), the transform can improve the overall performance by up to 2.5%. Among the five TFDs, CWT and CT achieved around 89.9% average MAcc, surpassing the others by approximately $0.5-1.3\%$ STFT and WVD achieved comparable MAcc of around 89.5%, slightly worse than CWT and CT. CWD performed the worst with 88.6%. Besides, it is worth noting that using a log-scaled raw signal can gain higher performance than the baseline raw signal, improving from 87.4% to 88.1% approximately.

Overall, CWT and CT were the two superior TFDs in our experiments, and there was no statistical difference between them ($p = 0.815$). STFT and WVD followed with trivial performance differences ($p = 0.413$). The p-values between CWT/CT and STFT/WVD were in the range of 0.012 to 0.035 in pair, indicating CWT/CT outperformed STFT/WVD statistically. CWD achieved slightly worse classification performance among the five TFDs, and its p-value concerning any other TFD was more petite than 0.001, indicating significant inferiority. Likewise, the baseline raw signal performed with the lowest MAcc, and its p-values concerning the others are all infinitely close to 0, indicating that using TFDs instead of the raw signal can improve the classification performance stably. Besides, the log-scaled raw signal achieved an intermediate performance between the baseline and the TFDs. Moreover, this can be proved objectively that its corresponding p-values approached 0 infinitely.



(a)

(b)

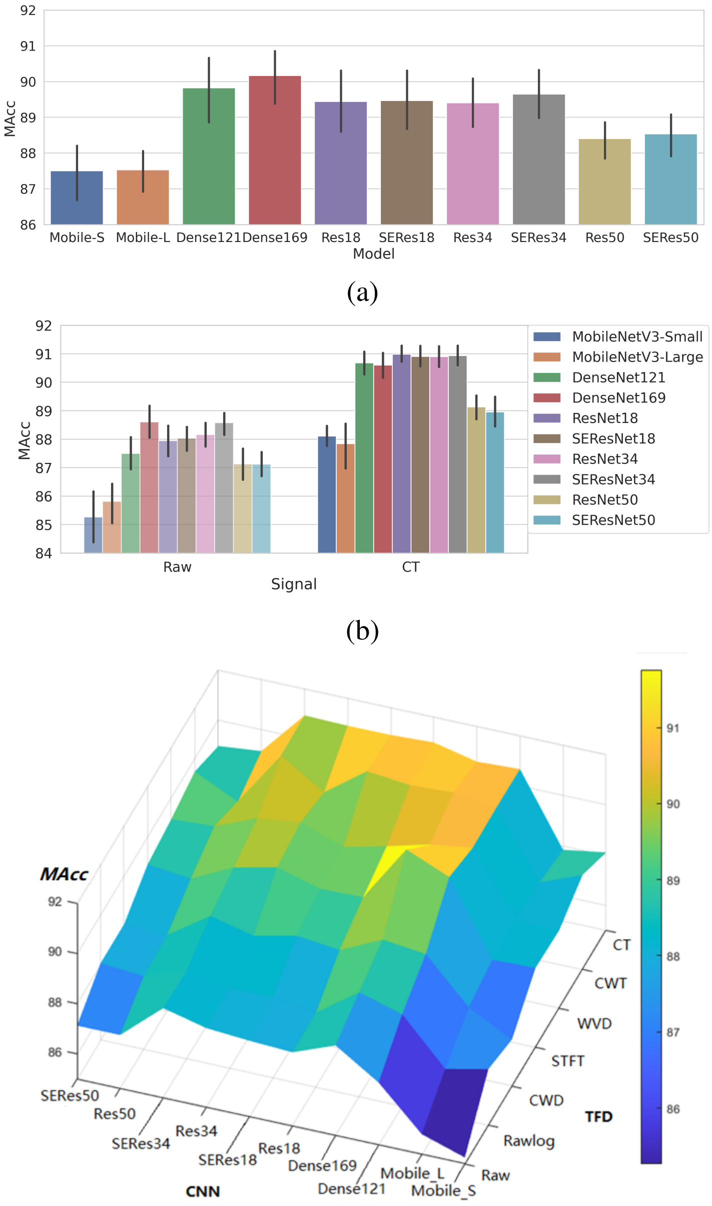
(c)

Fig. 5.7 (a) Overall performances of TFDs on all 10 CNNs, (b) Performances of TFDs on ResNet18, (c) Performances of TFDs on DenseNet121. The horizontal lines and the in-box white dots represent the median and average values, respectively.

As aforementioned, the overall performance differences were found among the TFDs. It is worth noting that the differences are CNN model dependent. For example, Fig. 5.7b and Fig. 5.7c demonstrated the performances of ResNet18 and DenseNet121, respectively. It can be observed that the difference ($> 1\%$) between STFT/WVD (MAcc: 89.47% / 89.96%) and CWT/CT (MAcc: 90.84% / 90.99%) existed for ResNet18. However, for DenseNet121, their performances were all around 90.8%, with differences smaller than 0.4%.

2) Comparison of the CNN performances

The overall performances (in MAcc, an average of 10 times) of 10 CNNs with all seven input types were illustrated in Fig. 5.8a. The models were sorted by the number of parameters from left to right. It can be observed that the lightweight MobileNets did not perform well, achieving approximately 87.50% MAcc. On the other hand, the performances of ResNet-18/34, SEResNet-18/34, and DenseNet121 were quite close, around 89.6% MAcc with differences below 0.4%. DenseNet169 outperformed all the others slightly and achieved 90.17% MAcc. Besides, ResNet50 and SEResNet50 using bottleneck block architecture achieved an unsatisfactory performance of around 88.5%.



(a)

(b)

(c)

Fig. 5.8 (a) Overall performances of 10 CNNs with all seven input types, (b) Performances of all the 10 CNNs on raw signal and CT. (c) Performances of the 10 CNNs concerning the seven TFDs.

To compare the performance of different CNNs, a statistical analysis was conducted. Particularly, the p-value between MobileNet-Small and MobileNet-Large was 0.791, which means they have no statistical difference. The p-values between ResNet-18/34 and SEResNet-18/34 were above 0.05 (in the range from 0.313 to 0.987), implying no statistical difference. The p-values of DenseNet121 concerning the four CNNs mentioned above ranged from 0.087 to 0.193, except for ResNet34 (0.031). This showed that DenseNet121 had a statically close performance with ResNet18 and SEResNet18/34 but was slightly better than ResNet34. DenseNet169 achieved relatively higher performance than the others, with p-values smaller than 0.017. Besides the CNNs mentioned above, ResNet50 and SEResNet50 shared close performance with a p-value of 0.357, higher than MobileNets but lower than the others.

Besides the overall performance of 10 CNNs, raw signal and CT performances were visualised on these CNNs as presented in Fig. 5.8b. The visualisation revealed the difference between raw signal and CT but presented different relationships between CNNs. For example, when using raw signal, DenseNet169 achieved the highest performance, 88.60%. But when using CT, it performed slightly worse than ResNet-18/34 or SEResNet-18/34 (DenseNet169: 90.60%, others: approximately 90.90%). Fig. 5.8c represents the overall performance using a surface plot.

3) Effect of combined TFDs as CNN Input

As described in 1), there was no apparent difference in the CNN performance using the TFDs except that CWD performed slightly worse. Since the TFDs transform the raw signals into different time-frequency domains, it is meaningful to analyse whether combining TFDs can improve CNN performances. Therefore, experiments were conducted on various channel-wise combinations of different TFDs as the inputs for CNNs. Notably, three CNNs representing three different network architectures with higher depth and better performance in 2) were selected, including ResNet34, SEResNet34 and DenseNet169, to analyze the effect of the TFD combinations. MobileNets were not included because of the slightly worse performance and shallower architectures had no advantage in fitting more information. Although SE/ResNet50 were deeper in architectures, they were not selected due to poorer performance compared with SE/ResNet34.

As aforementioned in 5.3.2, three out of the four TFDs (except CWD) are stacked in channel dimension to form the combined TFDs. Table 5.4 lists the performances (10 times average, sensitivity, specificity, and MAcc, respectively) of three CNNs using different single or combined TFDs. The results showed that combining input did not improve classification performance.

Table 5.4 Comparison of classification performance (%) between combined TFDs and single TFD (an average of 10 times).

Model	Combined vs. Single				Se	Sp	MAcc
	CT	CWT	STFT	WVD			
ResNet34	✓				85.83	95.97	90.90
		✓			83.85	95.97	89.80
			✓		84.17	95.65	89.91
				✓	84.81	95.49	90.15
	✓	✓	✓		85.51	95.24	90.38
	✓	✓		✓	86.11	95.37	90.74
	✓		✓	✓	83.44	94.94	89.19
		✓	✓	✓	85.80	95.41	90.61
SEResNet34	✓				85.61	96.27	90.94
		✓			85.86	95.91	90.89
			✓		83.50	95.56	89.53
				✓	84.33	95.56	89.94
	✓	✓	✓		86.50	95.35	90.93
	✓	✓		✓	86.31	95.91	91.11
	✓		✓	✓	83.85	95.32	89.59
		✓	✓	✓	84.52	95.50	90.01
DenseNet169	✓				85.76	95.44	90.60
		✓			86.11	95.73	90.92
			✓		**87.29**	96.22	**91.76**
				✓	85.25	95.56	90.41
	✓	✓	✓		85.92	**96.37**	91.14
	✓	✓		✓	85.57	96.29	90.93
	✓		✓	✓	85.89	95.79	90.84
		✓	✓	✓	86.66	96.31	91.48

5.4.4 Discussion

This study compared several commonly used TFDs to classify the heart sound signals on ten mainstream CNNs and explored using combined TFDs as inputs to improve the classification performance. In previous heart sound analysis studies, TFDs were only discussed for providing better human visualisation or more comprehensive information. They were used individually as CNN inputs in the classification, while no study has objectively compared and analysed their performance under different CNNs. Computer-aided analysis for heart sound is the future trend that shows the potential to overcome the physician's clinical experience and listening limitation, providing a more accurate and robust diagnosis. Thus, it is reasonably necessary to know if and how the selection of the TFDs will affect the CNN performance. This study filled this gap and researched the input and model selection. The results revealed that using log-scale for the raw signal as input in the time domain can improve classification performance with approximately $1-2\%$ in MAcc on MobileNetV3 but a trivial increase ($< 1\%$) on deeper networks. On almost all the models, converting the heart sound signal from the time domain into the time-frequency domain by TFDs can improve classification performance by approximately 2.3% in MAcc.

Furthermore, the selection of TFDs was not decisive on the heart sound classification, but it will still affect the performance. As illustrated in 1) of 5.3.3, where we compared all TFDs, CT and CWT are the two superior TFDs that outperformed STFT and WVD by a small margin and outperformed CWD. We assume that the slightly worse performance ($< 1\%$ than the rest TFDs) of CWD is due to the parameter setting ∂ of its kernel function. Although [200] claimed that $\partial = 3$ could provide better resolution, it may filter out useful information. Moreover, due to the different heart sound recording settings, a suitable ∂ may be further discussed. Even though TFDs were obtained by different mathematical calculations with different visual representations, as shown in Fig. 5.6, the hidden information provided was similar. Their morphological differences decrease significantly as CNN input with lower resolution (224×224), resulting in similar classification performances. Nevertheless, results suggest that CWT is more suitable than the rest TFDs as input for heart sound classification.

Secondly, the performances of different CNNs concerning TFDs were compared. It has been observed that the increase of model capacity was necessary for classifying the 2D inputs since the lightweight MobileNetV3-small/large performed worse than the other models with more parameters. However, only increasing the depth or the parameter number of CNNs without changing architecture does not guarantee a better result. This can be observed from ResNet and SEResNet family as in Fig. 5.8, that their classification performance remains the same when increasing depth from 18 to 34 (within 0.15% difference in MAcc, $p \leq 0.313$). Meanwhile, their performance significantly drops when increasing the depth from 34 to 50 for the two CNNs. This drop may come from the convolutional block. It changes from basic to bottleneck, where the convolutional features lose effectivity in deeper layers. SE was an approach to improve network performance, while there was no observed improvement in this study. This may be due to no extra valuable features were extracted by squeezing the convolutional features across the spatial dimension.

Furthermore, DenseNet-121/169 with fewer parameters achieve comparable performance to ResNet or SEResNet. It indicates that the dense layer-wise connection in DenseNet combining multi-scale convolutional features might benefit the heart sound classification. Based on the findings above, we conclude that increasing CNN capacity appropriately and extracting multi-scale convolutional features are two significant factors for improving CNNs performance in classifying 2D heart sound signals. Although MobileNetV3 achieved lower performance than the other more extensive networks ($< 3\%$ lower in MAcc), it costs much less computational resources as listed in Table 5.1. This indicates that some tiny networks like MobileNetV3 can perform heart sound classification in portable devices. In contrast, the other more extensive networks are more suitable for deploying in medical institutes or high-performance computing devices.

Thirdly, we explored the TFD combinations on three CNN models, including ResNet34, SEResNet34, and DenseNet169. The TFDs selected included CT, CWT, STFT, and WVD, while CWD was excluded due to its worse performance. The combination was implemented by placing three out of the four selected TFDs in the three channels of the input image for CNNs. The results in Table 5.4 indicate that the combination methods did not yield better performance than using single TFD for all selected CNNs. The lack of improvement can be explained by the fact that the TFD representations as CNN inputs were similar and sometimes overlapped; thus, the combined TFDs cannot provide extra helpful information than a single TFD.

Although the combination of TFDs did not improve the heart sound classification performance visibly, the performance of the CNNs and TFDs in this study is still considerable. Compared with rank 1 (0.8602 in MAcc) in the 2016 PhysioNet/ Computing in Cardiology (CinC) Challenge, the performance in this study (approximately 0.9 in MAcc) is higher. This showed that using the appropriate TFD and deep learning model is the optimum approach for heart sound classification.

This study revealed that when using CNN models to identify heart sound, it is beneficial to convert the raw signal from the time domain into the time-frequency domain, providing more comprehensive information to improve classification performance. Using TFDs can alleviate the need for training deeper CNNs to achieve identical performance. The effect of selecting time-frequency representations is not apparent for CNNs though they are different in human visualisation; we recommend CWT due to its stable performance in this study, generality, and understandability. For the CNN models, small-scale CNNs cannot perform as well as large-scale models (approximately $2-3\%$ less in MAcc), but they showed potential for embedded or portable devices. Among the mainstreamed large-scale CNN models, only increasing the depth or the parameter number is unnecessary for performance improvement and may even cause effectivity loss. The exploration of combining TFDs in the channel dimension as CNN input did not show improvement.

5.5 George B. Moody PhysioNet Challenges 2022

5.5.1 Introduction

Early screening is vital in detecting cardiovascular disease (CVD) and necessary action to reduce the risk of worsening heart disease. The initial suspicion often depends on the medical staff to listen to murmurs in the heart sound (recorded as phonocardiogram, PCG) during auscultation. However, due to the limitation of listening ability and clinical experience, auscultation is not always trustworthy [11]. Therefore, a more robust and accurate computer-aided PCG analysis algorithm is greatly needed to improve the situation.

The existing PCG classification methods can be divided into two types: (1) feature-based machine learning (ML) methods and (2) deep learning (DL) based methods. Feature-based ML requires manual extraction of the features, which heavily depends on PCG segmentation and feature settings. This usually causes robustness and portability issues. The inputs can be raw signals or their time-frequency distributions (TFDs) for DL-based approaches. Deep CNNs can extract the spatial features automatically, generally skip the segmentation and require fewer input settings. However, DL approaches require large datasets to improve classification performance. In recent years, large PCG datasets such as [42] and [232] have made DL approaches more competitive.

In the previous study [233], the 2-D TFDs as inputs for PCG classification were proved to outperform the raw signals on deep CNNs. Furthermore, the current mainstream CNNs were designed for the image recognition field with the local attention characteristic [234]. The receptive field of each CNN layer is fixed without considering the long short-term dependencies of the time-domain signal information. Therefore, the primary aim of this study is to design a novel CNN with hierarchical multi-scale architecture to improve the classification performance by fusing multi-scale features. In addition, the low-quality recording segments involving artefacts may mislead the network optimisation. Hence, the second aim of this study is to improve the classification accuracy by designing a quality assessment method to correct the labels for the low-quality segments. The proposed algorithm has been applied in the PhysioNet Challenge to verify the performance.

5.5.2 Methodology

1) Database and Pre-processing

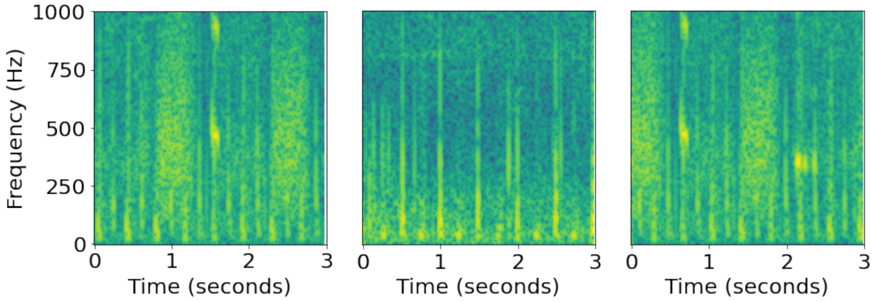
The database used in this study is the PhysioNet Challenge 2022 publicly released data, containing 3163 PCG recordings from 942 patients (Murmurs: 695 absent, 68 unknown and 179 present; Outcomes: 456 abnormal and 486 normal.). See [235] for more details.

In this study, the recording sampling frequency is downsampled from 4000 to 2000 Hz for faster data loading speed without information loss of the murmurs (ranging from $20-500$ Hz [236]). Afterwards, the signal is normalised by z-score normalisation.

2) Quality Assessment Method

The murmur labels given in the database correspond to each auscultation location recording of the patients, which means a recording is ’absent’, ’unknown’, or ’present’. In [229], the PCG duration effect has been proved minor on CNN performance. Thus, in this study, the recordings are cropped into 3s segments as CNN inputs, considering both the shortest signal length 5s and CNN receptive fields.

However, the PCG recordings contain many low-quality segments caused by ambient noise, artefacts, body friction, etc., which misleads CNN optimisation. Hence, a quality assessment for the murmur label correction of segments is needed. Since the frequency of normal heart sound is between $20-200$ Hz, the energy of most murmurs is much less than that of heart sound [236]. The selected assessment criteria is the ratio of spectral density between $20-200$ Hz to full band ($0-1000$ Hz), named quality ratio.



(a) Ratio 0.28 (b) Ratio 0.52 (c) Ratio 0.26

Fig. 5.9 Spectrogram of a segment with a quality ratio of (a) 0.28 from signal labelled ’absent’. (b) 0.52 from the same signal. (c) 0.26 from signal labelled ’unknown’.

Fig. 5.9a and 5.9b are spectrograms of two segments from one ’absent’ recording. Fig. 5.9c is of a segment from an ’unknown’ recording. There are visible differences between 5.9a and 5.9b, especially in the higher frequency bands. On the contrary, these high-frequency noises in 5.9a are similar to those in 5.9c. After manual frequency analysis on recordings, the label correction strategy is: if the quality ratio is larger than 30%, this segment murmur label follows the recording label. Otherwise, it is relabelled as ’unknown’. It should be noted this label correction is only for murmur labels but keep outcome labels unchanged.

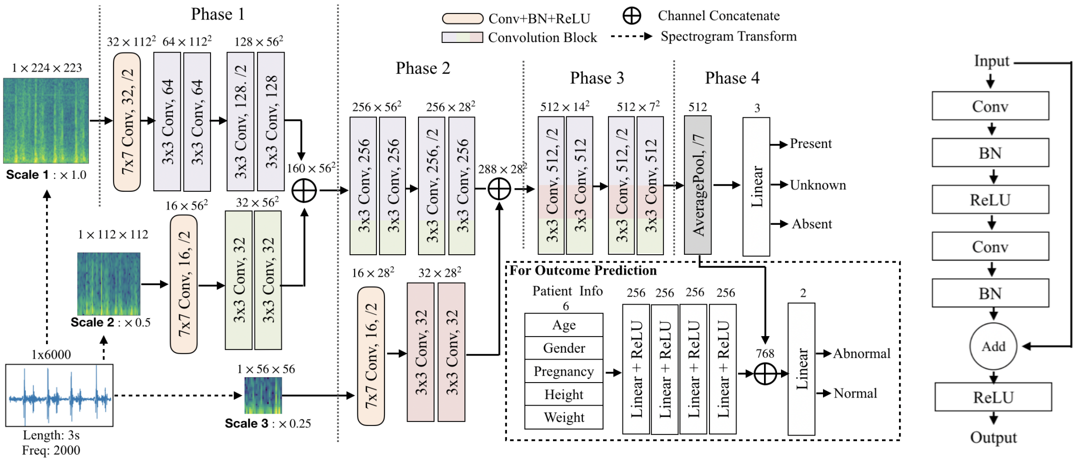
3) Model Interpretation

The CNN inputs are the multi-scale spectrograms of 3s segments. Three scales ($\times 1.0$, $\times 0.5$, $\times 0.25$) are selected. The parameters for the spectrograms are given in Table 5.5.

The multi-scale spectrograms provide CNNs with time-frequency features in different resolutions and reduce the spatial information loss in single spectrogram.

Table 5.5 Parameter settings for multi-scale spectrogram.

Scale	Nfft	Window length	Hop length
×1.0	446	200	27
×0.5	222	100	54
×0.25	110	50	108



(a) (b)

Fig. 5.10 (a) Overall structure of the HMS-Net. The in-box text denotes the layer parameters, e.g., '3 × 3 Conv, 128, /2' represents a convolution layer with 128 3 × 3 kernels and stride 2. The text above boxes denotes the output size. The colours in convolution blocks indicate the information from certain scales. (b) Structure of a residual convolution block.

Inspired by [237, 238], in this study, a hierarchical multi-scale convolutional neural network (HMS-Net) is proposed to improve the PCG classification performance by building long short-term dependencies between multi-scale inputs with its hierarchical architecture. Fig. 5.10a illustrates its overall structure. The fundamental element in HMS-Net, convolutional block, refers to ResNet [205], with its structure diagram shown in Fig. 5.10b. Three-scale spectrograms of a segment input HMS-Net and output the 3-class murmur prediction. For outcomes prediction, it combines with patient information and outputs the binary result.

In HMS-Net, the convolutional features of the multi-scale spectrograms are extracted at different depths. A larger scale requires deeper layers; thus, HMS-Net has four phases containing layers with incremental depths for extracting features from different scales. For example, in Phase 1, two sub-networks are employed to convolve the features from Scale 1 and Scale 2. The 2-scale features are then concatenated in channel dimension and passed to the next phase. Phase 4 summarises the multi-scale convolutional features by global average pooling and classifies the segment with a linear layer. Overall, HMS-Net extracts features from multiple scales separately at the beginning and fuse these features with its hierarchical design.

Regarding the outcome classifier, patient information, including age, gender (one-hot), pregnant status, height, and weight, is added as extra information to distinguish patients with abnormal clinical outcomes. As shown in Fig. 5.10a, 256 patient features are extracted from these information via a 4-layer multi-layer perceptron (MLP). The final outcome prediction is obtained from both the convolutional features and the patient features.

4) Training Settings

The optimiser is Adamw and the max training epoch is set to 100. The initial learning rate is 10^{-3}. When the training loss has stopped decreasing for five epochs, the learning rate is multiplied by 0.1. The loss function is cross-entropy with 0.1 label smoothing. In each batch, 128 multi-scale spectrograms are fed to HMS-Net.

5) Murmur Classification

Since the HMS-Net is designed to classify the segments, for recording classification, a sliding window with 3s width and 1s step is applied to classify the whole recording continuously. For a frame slided by multiple windows, its label is calculated by the averaged distribution probabilities of the passed windows. The prediction for the recording is the serial labels per second. If the predicted ’unknown’ accounts for over 80% of the serial labels, the recording is catalogued as ’unknown’. Otherwise, the recording is classified with the majority of serial labels (exclude ’unknown’).

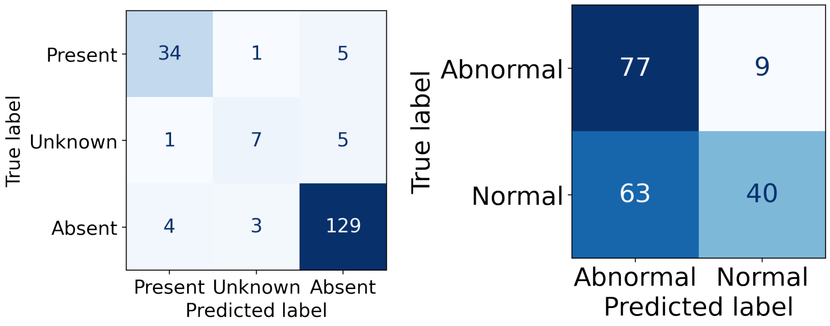
For patient classification, if one location recording is classified as ’present’, the patient is labelled ’present’. In terms of ’absent’ and ’unknown’, the patients are classified by the majority of location recording labels. When there is the same number of ’absent’ and ’unknown’ recordings of the patient, ’absent’ has the priority. All the mentioned thresholds are chosen based on local testing.

6) Outcome Classification

Our outcome prediction strategy is similar to murmur prediction but does not involve ’unknown’ issue. The serial labels (’abnormal’ or ’normal’) per second are obtained by sliding window as well. If over $1/3$ frames are predicted as ’abnormal’, the recording is predicted as ’abnormal’. When a patient has at least one predicted ’abnormal’ recording, our strategy diagnoses the patient as ’abnormal’. Otherwise, the patient is diagnosed as ’normal’.

5.5.3 Results

We used five-fold cross-validation by patients to fairly evaluate our methods. See [235] for the scoring metrics of murmur weighted accuracy and outcome challenge score.



(a) (b)

Fig. 5.11 Confusion matrices of (a) murmur classification (b) outcome classification.

Our method achieved an average murmur classification accuracy of 91.37% (best 92.85%) on segments in the five-fold cross-validation. It performed 83.78% averaged murmur classification accuracy on patients and 0.81 averaged weighted score. Fig. 5.11a shows the confusion matrix of the best fold on patient classification. The overall accuracy was 89.94%, respectively, on 'present' was 85.0%, 'unknown' was 53.84% and 'absent' was 94.85%. The weighted murmur accuracy was 0.853. Regarding patient outcome classification, our method achieved averaged 56.83% accuracy (best 62.96%) and 9808 averaged outcome (best 9242). The outcome confusion matrix for highest outcome score is shown in Fig. 5.11b. In the blind validation set, the algorithm achieved 0.806 murmur weighted accuracy and 9120 outcome challenge score, and in the blind testing set, they were 0.776 and 12069, respectively.

5.5.4 Discussion

Label Correction The low-quality ('unknown') segments or recordings caused by artefacts exist considerably and are often fused with heart sounds and murmurs. When training the CNN model on segment inputs, these low-quality parts greatly mislead the model. It is necessary to identify them while often being neglected. Therefore, a quality assessment method by spectral density was proposed to alleviate the label inconsistency problem. With the assessment method, the accuracy on segments increased by approximately 5%. Though, in the current method design, it did not involve too many criteria considering the data loading speed for the DL methods. There is huge room to extend the assessment criteria by the PCG time-domain or frequency-domain features to achieve better label correction and improve the classification accuracy. In future work, removing the unknown segments in data pre-processing or alleviating its effect in the testing interface should be studied.

HMS-Net HMS-Net holds the advantage of combining the features from multi-scale spectrograms to improve the classification performance. However, work could still be done on determining the optimal network depth and width, parameter optimisation, etc., to make the network more efficient. Furthermore, the low-quality segments issue made it hard to objectively evaluate its segment classification performance when many label inconsistencies occurred. This is also why the confusion matrix on segments was not provided in the results. Despite this, in local tests, HMS-Net performed approximately 1% better than ResNet34.

Outcome Prediction The clinical outcomes diagnosed by cardiologists are based on multiple assessments. Only PCG with basic patient information is far from enough to reliably and accurately identify the outcome. More patient diagnostic information like echocardiogram can be served as extra inputs to provide CNNs with more valuable information. Besides, the outcome result is quite sensitive to hyper-parameter settings.

Overall, this study proposed a hierarchical multi-scale convolutional neural network with a signal quality assessment method to classify PCG. In the PhysioNet Challenge 2022, it performed outstandingly with 1^{st} in murmur detection and also top scores in clinical outcome. The proposed method may be inspiring and significant in future PCG classification design.

5.6 Conclusion

This chapter discusses the works conducted on PCG classification. In Section 5.3, we analyze the effect of signal duration on the classification of heart sounds. The results of this study revealed that (1) very short heart sound signal duration (1 s) weakens the performance of RNNs, whereas no apparent decrease in the tested CNN model was found. (2) RNN outperformed CNN using Mel-frequency cepstrum coefficients (MFCCs) as features. There was no difference between RNN models (LSTM, BiLSTM, GRU, or BiGRU). (3) Adding dynamic information ($\Delta\&\Delta^2 MFCCs$)) of the heart sound as a feature did not improve the RNNs' performance, and the improvement on CNN was also minimal ($\leq$ 2.5% in MAcc). The findings provided a theoretical basis for further heart sound classification using deep learning techniques when selecting the input length.

In Section 5.4, we investigate the optimal use of TFD/ combined TFDs as input for CNNs. The presented results revealed that: (1) The transformation of the heart sound signal into the TF domain achieves higher classification performance than using of raw signals. Among the TFDs, the difference in the performance was slight for all the CNN models (within 1.3% in average accuracy). However, Continuous wavelet transform (CWT) and Chirplet transform (CT) outperformed the rest. (2) The appropriate increase of the CNN capacity and architecture optimisation can improve the performance, while the network architecture should not be overly complicated. Based on the ResNet or SEResNet family results, the increase in the number of parameters and the depth of the structure do not improve the performance apparently. (3) Combining TFDs as CNN inputs did not significantly improve the classification results. The findings of this study provided the knowledge for selecting TFDs as CNN input and designing CNN architecture for heart sound classification.

In Section 5.5, we proposed the wining algorithm with a recording quality assessment method based on frequency density distribution for label correction to prevent the poor-quality recording segments from misleading network optimisation. Besides, a hierarchical multi-scale convolutional neural network (HMS-Net) was designed to conduct both the murmur (T1) and clinical outcome (T2) classification. HMS-Net extracts convolutional features from the spectrograms on multiple scales and fuses them through its hierarchical architecture. The network builds long short-term independencies between multi-scale features and improves the classification performance. Finally, the prediction of a patient is based on the ensembled segment predictions by sliding window. In the five-fold cross-validation by patients, the proposed algorithm performed an average weighted accuracy of 0.81 (best 0.853) on T1 and an average challenge score of 9808 (best 9242) on T2. In the Challenge hidden validation set, the proposed algorithm achieved 0.806 weighted accuracy on T1 and 9120 challenge score on T2, ranking 1^{st} and 4^{th} out of 305 entries, respectively. In the final hidden testing set, T1 was 0.776 ranking 1^{st}, and T2 was 12069.

Chapter 6

Classification of Paroxysmal Atrial Fibrillation using Deep Learning Methods on ECG Signals

6.1 Introduction

Atrial fibrillation (AFib) is an irregular heartbeat (arrhythmia) caused by the ectopic impulses in the atrium. It may lead to blood clots, stroke, and heart failure, which are severe hidden dangers to human lives. Furthermore, the AFib is a common issue for approximately 2% of people younger than 65 and 9% older than 65 [239]. The American Heart Association guideline [240] classified Afib into four types: paroxysmal AFib, persistent AFib, long-standing persistent AFib, and permanent AFib based on the duration and recoverability. While in clinics, physicians usually sort them into paroxysmal and persistent types only. Paroxysmal AFib episodes can last several seconds, hours, or even days before returning to normal sinus rhythm. Lack of intervention may lead the paroxysmal into persistent AFib, which is irreversible. Due to the intermittent characteristics of the paroxysmal AFib, it is generally neglected by patients before deteriorating into a persistent type. As a result, the all-cause mortality rate is approximately 6.3% on AFib patients [241]. Therefore, it is vital to have an algorithm that can work automatically in the early screening to prevent the paroxysmal AFib from worsening to persistent AFib or more severe health issues.

ECG is the most commonly used approach in cardiac diagnosis. It represents the electrical activity of the heart. The whole electrical process starts with the spontaneous impulse generated at the Sinoatrial node (SA node), then propagates to the atrioventricular node (AV node), causing the squeezing of the atria as represented by the P wave. Afterwards, the electrical signal is transmitted to the His bundle and Purkinje fibres, causing the contraction of the ventricles. The ventricles repolarizes and ready for the next heart cycle.

The QRS complex indicates the depolarization, and the T wave shows the repolarization of the ventricles, respectively. However, AFib is caused by irregular fast squeezing of the atria leading the heart walls quiver, or fibrillate. This phenomenon it is reflected by disorganized electrical activity (ectopic impulses instead of SA impulse) in the atrium, so its ECG signal differs from normal, as shown in Figure 6.1. Morphologically, the AFib ECG has irregular intervals, a narrow QRS complex, and undulating P waves. Thus, using ECG signals to identify the AFib is a practical approach in designing automatic classification algorithms.

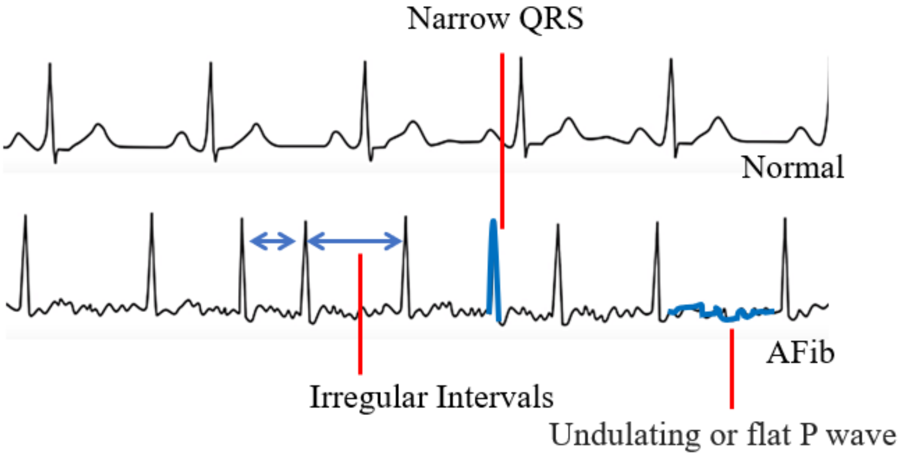


Fig. 6.1 The cardiac cycles of normal and AFib ECG.

Computer-aided algorithms for AFib detection have been developed for decades, and the proposed algorithms covered the conventional machine learning (ML) methods such as support vector machine (SVM), k-nearest neighbours algorithm (KNN), random forest, discriminant analysis, etc [242–247]. These conventional ML approaches relied on manually extracted features such as average, standard deviation, and entropy of RR intervals in the time domain [248], power spectral density in the frequency domain, and statistical features such as kurtosis and skewness [247]. With the development of deep learning (DL) in recent years, approaches such as convolutional neural network (CNN) and recurrent neural network (RNN) have also been tested on AFib detection [249–251]. They hold the advantage of neglecting feature extraction and using raw ECG signals as input and have also achieved promising performance. Though there are tons of researches focusing on AFib classification, only few pieces of research work have focused on paroxysmal AFib detection due to the lack of suitable databases. As a result, paroxysmal AFib is often unrecognized [252]. Therefore, it is pretty meaningful to explore the capability of the neural network (NN) in the identification of paroxysmal AFib.

In this study, the primary aim is to propose an algorithm that can classify the non-AFib, persistent AFib, paroxysmal AFib, and their onsets. The secondary task is to constrain the computing load while achieving comparable performance, making it available for a standard laptop or embedded system. All the findings provide knowledge on using NNs to classify paroxysmal AFib and contribute to designing small-scale portable ECG devices which can do real-time monitoring of the heart conditions.

6.2 Two-Stages RNN Algorithm with Feature Extraction

6.2.1 Database and Pre-processing

The database used in this research was CPSC2021[253]. It includes 1436 ECG recordings (475 Persistent AFib, 229 Paroxysmal AFib, 732 Non-AFib) from 100 subjects (24 Persistent AFib, 23 Paroxysmal AFib, 53 Non-AFib).

In this study, a two-stage algorithm is designed to conduct the detection of paroxysmal AFib and its onsets. The flowchart of the proposed algorithm is shown in Figure 6.2. In Stage I, a Bidirectional Long short-term memory (BiLSTM) network was used to classify the ECG segments into Non-AFib and AFib segments. Then the ECG signals consisting of AFib segments were transferred to Stage II and classified into Persistent AFib or Paroxysmal AFib. A moving window was employed to classify the whole signal and detect AFib onsets. The processing was conducted in Matlab® R2021a environment, using a laptop (CPU: i7-8650U, RAM: 16G, no GPU).

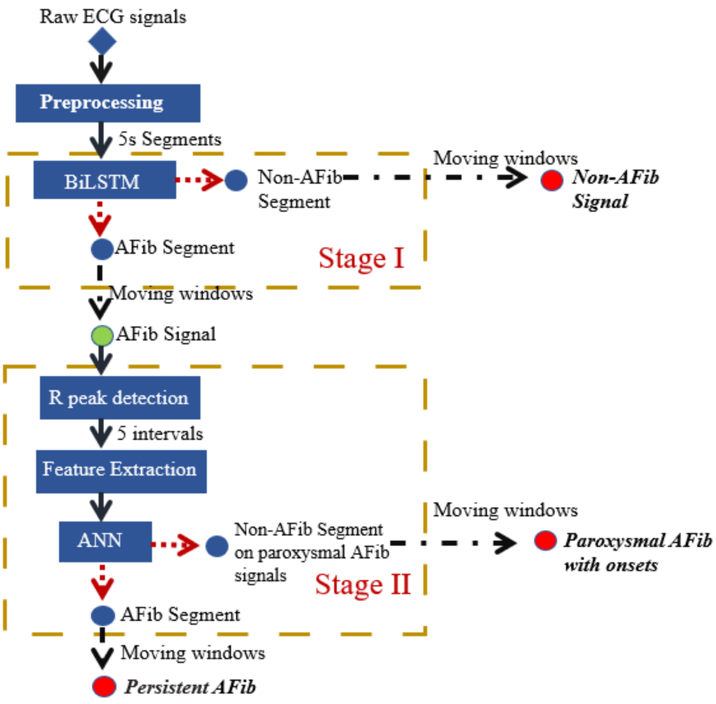


Fig. 6.2 The flowchart of the designed algorithm.

Before the two classification stages, the ECG signals were pre-processed. The raw ECG signals were normalized (z-score), filtered with 0.5 – 30 Hz bandpass filter (3^{rd} order Butterworth), then segmented into 5s segments for training (without overlap). After segmentation, 699040 ECG segments were generated (421022 Non-AFib, 212098 Persistent, and 65920 Paroxysmal) for training.

6.2.2 Evaluation Metrics

The validation accuracy of the two stages indicates their capability to identify the small segments (within windows). The overall performance of the algorithm can be reflected by the score of the testing recordings. In this paper, the CSPC2021 Challenge scoring scheme is considered [253].

The score includes two parts: the first part (Ur) classifies the AFib correctly, and the score matrix is shown in Figure 6.3. The second part (Ue) is meant to detect the AFib onsets. If the onsets and end of the AFib episodes were detected within ±1 R-peak, $Ue+1$, within ±2, $Ue+0.5$.

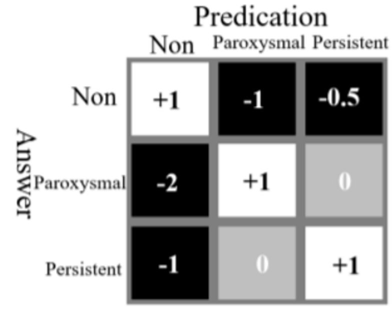


Fig. 6.3 The score matrix for part one.

The overall score (U) is calculated by:

$$U = \frac{1}{N} \sum_{i=1}^{N} \left(Ur_i + \frac{Ma_i}{\max\{Mr_i, Ma_i\}} \times Ue_i \right), \tag{6.1}$$

6.2.3 Architecture of the Algorithm

Stage I: BiLSTM

BiLSTM is one type of RNN algorithms that showed outstanding performance in the analysis of data sequence, such as speech and text recognition [254, 185]. In the proposed algorithm, a simplified structure with two layers of BiLSTM (hidden units: 50) was applied. The inputs for the BiLSTM layers were 5s segments. After BiLSTM, it is connected with a fully connected layer to project the results into Non-AFib (0) and AFib (1) two classes. The overall structure of Stage I is shown in Figure 6.4(a).

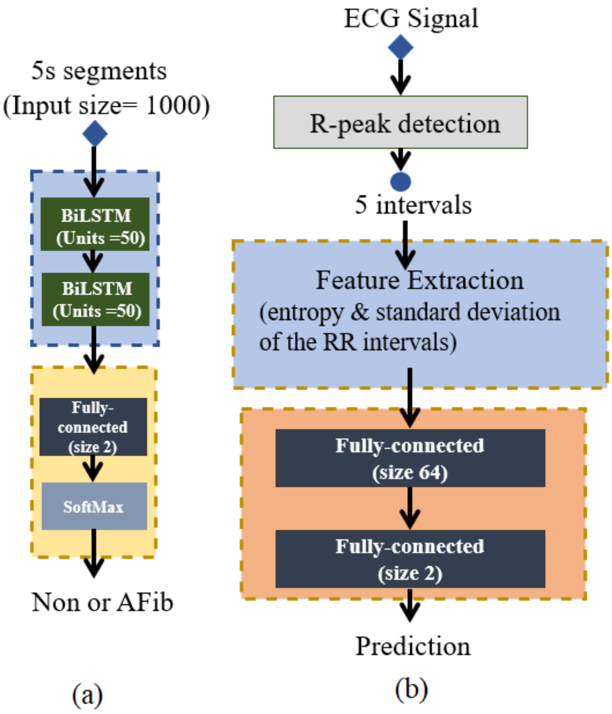


Fig. 6.4 (a) The Stage I structure, (b) The Stage II structure.

During training, the used training sets included non-AFib segments labeled (0), persistent AFib segments labeled (1), while paroxysmal AFib segments were also labeled (1) to increase the sensitivity. Training and Validation Proportion was 7:3. 20% recordings (285) were randomly left for the whole signal testing, including 145 non-AFib, 95 persistent, and 45 paroxysmal recordings. The optimizer selected in this study was stochastic gradient descent with momentum (SGDM). The initial learning rate was 0.001 with a drop factor of 0.2, the max epoch of 10, and the batch size of 256. The network can identify the non-AFib segments of the ECG signal. For the complete signal classification, a moving window (size: 5s, slide: 1s) was conducted on the signal to classify each segment. A majority voting was applied to avoid sudden incorrect classification. Each time frame was covered by 5 sliding windows, so the time frame is only labeled AFib when more than 3 windows (segments) were classified as AFib.

Stage II: Feature extraction ANN

In the testing phase, the use of a relatively simplified DL network (with two layers BiLSTM and three Conventional layers structure, such as Stage I) didn't perform well in the identification of paroxysmal or persistent AFib. The loss didn't go down and the training accuracy remained at 69.15%, which means the network was uncapable to learn.

Deeper and complex network structure were excluded to avoid increasing the computation burden. Therefore, manual features extraction was applied in the classification stage, where entropy and standard deviation of RR intervals (which are commonly used as input features for classification) were selected. The process of Stage II is shown in Figure 6.4(b). R-peaks were extracted by Pan–Tompkins algorithm (Pan and Tompkins, 1985). Five RR intervals were clipped as a segment, and the entropy and standard deviation were extracted from the segments. Afterward, they were sent to the fully connected layers to classify into non-AFib or AFib segments. Similar to Stage I, a moving window (size: 5 intervals, slide: 1 interval) was also applied to identify the whole signal as persistent or paroxysmal. The entropy calculation is given by the equation:

$$\mathrm{E}(R) = -\sum_{i=1}^{n} \mathrm{P}\left(R_i\right) \log \mathrm{P}\left(R_i\right) \tag{6.2}$$

where E is the entropy of the segment, R_i indicates each RR interval length and P is the occurrence probability.

The training sets were only persistent AFib labelled (1), and paroxysmal AFib segments were labelled according to the reference label. Because the paroxysmal segments were approximately 30% of persistent segments, and the non-AFib segments of the paroxysmal are less. Therefore, a moving window (size: 5 intervals, slide: 1 interval) was applied to section more paroxysmal segments to balance the data structure. The rest training settings were the same as Stage I.

6.2.4 Results and Discussion

For Stage I, the validation sets achieved 90.14% accuracy to classify the non-AFib and AFib segments with a specificity of 93.65% and sensitivity of 84.82%, respectively. The result indicated that Stage I could identify the non-AFib segments well but may miss some AFib segments. However, it wasn't an issue for the whole signal because the majority voting and the appropriate threshold can improve the overall performance and remedy the sensitivity. In the testing phase, a 2.5% threshold was set which means if less than 2.5% of the signal is classified as AFib, the overall signal is regarded as non-AFib. By this approach, the accuracy of non-AFib signals classification could be increased approximately from 92.62% to 96%. Theoretically, raising the threshold can improve the non-AFib accuracy on validation to almost 100%, but it loses its sensitivity and generalization.

For Stage II, on the validation sets, it achieved the accuracy of 92.56% with a specificity of 86.24% and sensitivity of 95.77% to classify non-AFib and AFib segments on the AFib signals. The result showed that Stage II might tend to classify the healthy segments into AFib segments. However, because of the considered two stages design, non-AFib signals have been excluded before Stage II; thus, it won't affect the overall classification performance. It only affects the detection of the onset of the AFib.

During the testing recordings, the two-stages method achieved 2.0953 overall mark, including 0.8714 Ur and 1.4039 Ue. It showed a satisfying performance on the classification, while the onset detection can be improved. Furthermore, the total neural network is only about 1.6 MB in Matlab (coding in Python can be smaller, approximately 500 k.), which is possible to use on a personal laptop or embedded device.

This study aimed to design an algorithm using NNs to detect paroxysmal AFib and make the computing load small enough for a portable embedded ECG device. This is done because patients typically neglect paroxysmal AFib due to its intermittent characteristics and lack of appropriate databases. In this study, a two-stage algorithm was designed using the CSCP2021 database, and its capability to classify the AFib segments and onsets on the validation sets proved. Firstly, the use of a two-stage method rather than one NN is justified. Before the training, our preconceived thought on paroxysmal AFib was like intermittent non-AFib and AFib waveforms in the ECG signals. However, it is not, or at least the BiLSTM or Conventional Neural Network (CNN) cannot easily learn it. For non-AFib or AFib segments from the non/persistent AFib signals, the network in Stage I can learn in a very short time within one epoch, while the segments from paroxysmal could not regress, and the loss didn't go down (training accuracy also stuck at 69.15%, which is approximately equal to the data proportion). This may indicate that the paroxysmal AFib may hold pathological characteristics even in the healthy episodes and using a simplified network cannot classify the non-AFib or AFib episodes. There is no doubt that using the deeper NN with complex structure, such as adding lots of CNN layers and attention layers, can learn the difference. Still, it makes the computing load quite extensive, which is contrary to the original intention. Therefore, a second phase was included for the detection of the paroxysmal onset. Secondly, the use of Stage II to finish the whole classification task is tested. However, the performance was not satisfying due to the oversensitivity of the Stage II network and its trend to identify the segment as AFib. Besides, feature extraction relies greatly on reliable and accurate R peak detection. When the signal has massive motion artefacts, the failed R peak detection causes an error in the algorithm. This is another advantage of the two-stage structure. Thirdly, there is still room for the improvement of the overall performance. In the blind test of the challenge, the overall mark is decreased from 2 to approximately 1.7. This result showed that the generalization needs to be improved, especially in Stage II. Currently, only two features were used while adding more features might be a solution to improve the algorithm. Besides, appropriate window length may also affect the result. Currently, a 5s window on Stage I and five intervals on Stage II are used. Longer window length may provide more information, especially on the feature extraction of Stage II. Short duration cannot maximize the feature difference.

6.3 Conclusion

This study proposed a two-stage neural network algorithm that can detect paroxysmal AFib and its onsets. For performance, it can achieve 90.14% and 92.56% accuracy on non-AFib and AFib segments classification respectively in the two stages, got 2.0953 overall mark on our testing sets. As few researches have focused on paroxysmal AFib detection using NNs, the finding of this study provides knowledge for the further researches in this area. In the meantime, the proposed method also holds the advantage of a small computing load, making it possible for embedded ECG devices.

Chapter 7

Studies on EDR Extraction

7.1 Introduction

The respiratory rate (RR) is a vital physiological parameter in prediagnosis and daily monitoring. It can be obtained indirectly from Electrocardiogram (ECG) signals using ECG-derived respiration (EDR) techniques. As part of the study in designing an early cardiac screening system, this part of thesis aimes to provide accurate respiratory rate while eliminating the need of physical respiration sensor using EDR methods.

This chapter introduces the three conducted studies on EDR extraction. Section 7.2 is to assess the feasibility of extracting EDR from the localised ECG at the auscultation sites by experiments. Section 7.3 is about exploring using embroidered electrodes to extract the EDR. Section 7.4 introduces the progress on EDR under stress test. Section 7.5 summarizes the works.

7.2 Estimation of the Respiratory Rate from Localised ECG at Different Auscultation Sites

7.2.1 Introduction

Since the researches on EDR mainly focused on traditional 12-leads ECG and barely on localized ECG signal. Therefore, it is worth researching how the EDR performance within auscultation sites is. The aim of this study is threefold: (1) To investigate if the location of the electrodes at auscultation sites affect the EDR algorithm accuracy; (2) to compare the performance of one-lead EDR algorithms based on the mentioned respiration-induced ECG variation; (3) to compare time-domain and frequency-domain features for RR estimation. All the findings will contribute to providing more accurate RRs for the integrated cardiac screening device.

7.2.2 Methodology

1) Subjects

The experiments were conducted on 12 healthy human subjects (8 male/4 female, age range 21-29 years, mean 25.9 years) with no history of heart diseases or respiratory issues. The procedures were approved by the King's College Research Ethics Committee (Approval No.: LRS-18/19-10673). Subjects gave written informed consent before the experimental procedures.

2) Experimental Setup

The standard Lead I ECG (as reference ECG), auscultation site ECG (captured at auscultation site A, P, T, M with 10 cm inter-electrode distance), and respiratory signals were recorded simultaneously during the experiment. A simple block diagram of the experimental setup is shown in Fig. 7.1. The sensors for ECG signals were solid gel electrodes (Ambu WS, size: 36 × 40 mm, Medico Electrodes International LTD., Uttar Pradesh, India), and the respiratory signal was captured by a small microphone (developed at the Centre for Robotics Research (CORE) at Kings College London, UK) placed under the subject's nose. The recording used the commercial acquisition system (iWorx, model RA834, iWorx Systems Inc, Dover, NH, US) and ECG devices (iWire-BIO4, iWorx Systems Inc, Dover, New Hampshire, US). The sampling frequency was 1 kHz and the analog filter for the ECG was 0.05–40 Hz [177].

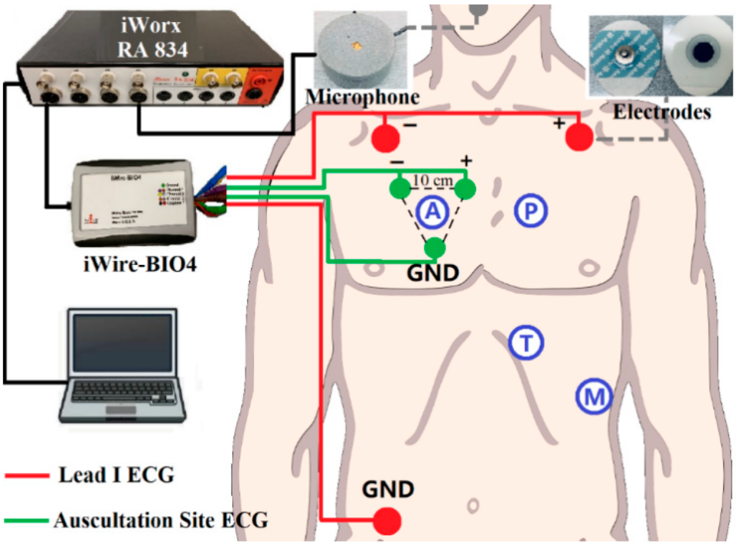


Fig. 7.1 Block diagram of the recording setup: Red dots are Lead I Electrocardiogram (ECG) as a reference, green dots are auscultation site ECG. The grey dot is the microphone for respiration recording. iWire BIO4 is for ECG recording. All the data is transferred to the computer for processing through iWorx RA 834.

During the experiment, subjects should keep supine and remain calm. Besides, subjects were required not to make sound from the larynx to ensure the sound captured was only respiration. The Lead I ECG and different auscultation site ECG signals were measured in pairs together with the respiratory signals. The duration of each recording group was three minutes, and two minutes break was given between different auscultation site trials.

3) Signal Processing

In this study, EDR signals were obtained using BW, AM, FM, and BP algorithms from the reference (Lead I) and auscultation sites ECG signals, respectively. The RRs were estimated from the EDR signals using time and frequency domain features as detailed later. The performance of the algorithms and the effect of the locations were analysed by comparing it with the measured respiratory rate. The processing was conducted in the Matlab ® $R2018b$ environment, and the statistical analysis was performed using IBM ® SPSS version 26.

The captured ECG signals and respiration sounds were filtered first to remove the unwanted artifacts and noise. For the ECG, a zero-phase 3^{rd}-order Butterworth high-pass filter at 0.1 Hz was used to eliminate the large artifacts which were not related to respiration [178]. For the respiration sound, a 3^{rd}-order Butterworth band-pass filtered (0.1–0.5 Hz) was used to smooth the waveform.

In AM, BW, and FM algorithms, R-peak detection was a vital step, as all the features to be captured were related to R peaks. In this study, the Pan–Tompkins algorithm was used to detect R-peaks in the ECG signals [89].

1. AM algorithm: The amplitude changes due to the respiration in the ECG signals was obtained by connecting the captured R-peaks.

2. BW algorithm: Based on the R-peaks, Q points were found using the gradient descent method. Then, the baseline wander could be generated by connecting the middle points between R-peaks and Q points [144].

3. FM algorithm: The intervals between the R peaks were calculated. The resulting signal was the frequency modulation caused by respiratory sinus arrhythmia. Afterward, all the signals generated by the algorithms above were interpolated to the same sample size of its raw ECG signals to increase the resolution.

4. BP algorithm: A band-pass filter (0.1–0.5 Hz) was used to capture the EDR signals. Although the normal RR for a healthy adult ranges between 0.2–0.35 Hz at rest, in our processing, we appropriately expanded the range to enable it to respond to special situations, such as the subjects' occasional deep or rapid breaths. Besides, a wider band can help to further analyse the frequency components when there are no dominant peaks. Representative derived respiration signals by the methods above are shown in Fig. 7.2.

The reference RRs were obtained from the filtered respiration sound recorded using a nostril microphone. It was manually counted in the waveform to ensure accuracy. The

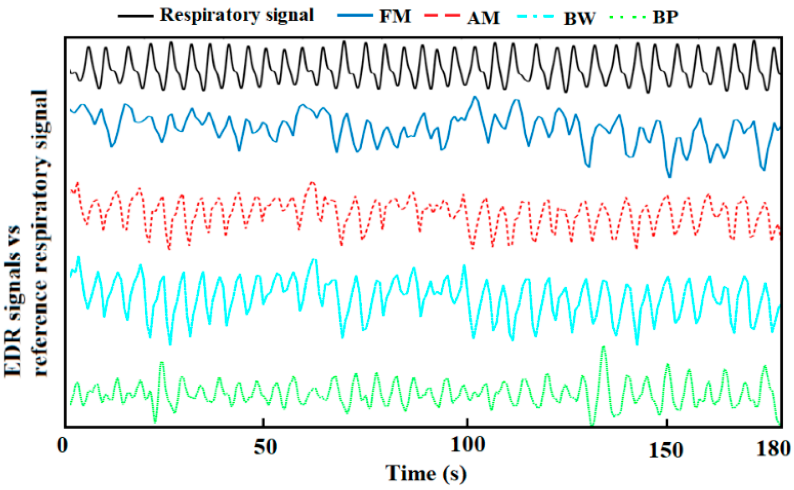


Fig. 7.2 A representative derived respiration signals from auscultation site ECG and reference respiration signal.

estimated RRs from EDR signals were calculated by automatically counting in the time domain and using the median frequency (between 0.1–0.5 Hz), respectively. For the counting method, a moving average filter (window length: 50 ms) was used first to smooth the EDR signals and eliminate sub-peaks. Then, peak detection with the threshold of the signal mean value provided the estimated RR. The median frequency was chosen according to our previous study, which was proven to be the best feature in the frequency domain to estimate RR from EDR signals [162].

4) Statistical Analysis

The mean absolute errors (MAE) between the EDR-based estimated RR and reference RRs was used as the performance measure provided as mean ± standard error (SE). A three-way repeated-measures analysis of variance (ANOVA) was used to compare MAE. Factors were the features (counting and median frequency), EDR algorithms (AM, BW, FM, BP), and ECG locations (A, P, T, M, Lead I). A P-value of less than 0.05 was considered significant. Data were log-transformed to obey normality and variance homogeneity was satisfied.

7.2.3 Results

1) ECG Morphological Variation among the Auscultation Sites

Fig. 7.3 shows a representative local ECG morphological variation compared with the lead I ECG from one subject. From (a) to (d), it can be seen that the amplitude of the R-peak, s-wave, and T-wave become larger from auscultation site A to M. Besides, it is also found that the R-peak of the site A ECG is normally on the left-hand side of it on Lead I ECG, which means the R-peak is advanced (approximately 10 ms by average). However, it shifts to the right-hand side when measured at site M, which means its onset is delayed (approximately 15 ms by average). Another phenomenon could also be observed that in the site A ECG, there is a J-point elevation shown as grey dots in (e). This happened on five subjects, and in 3 of them, the J point is even higher than the R-peak.

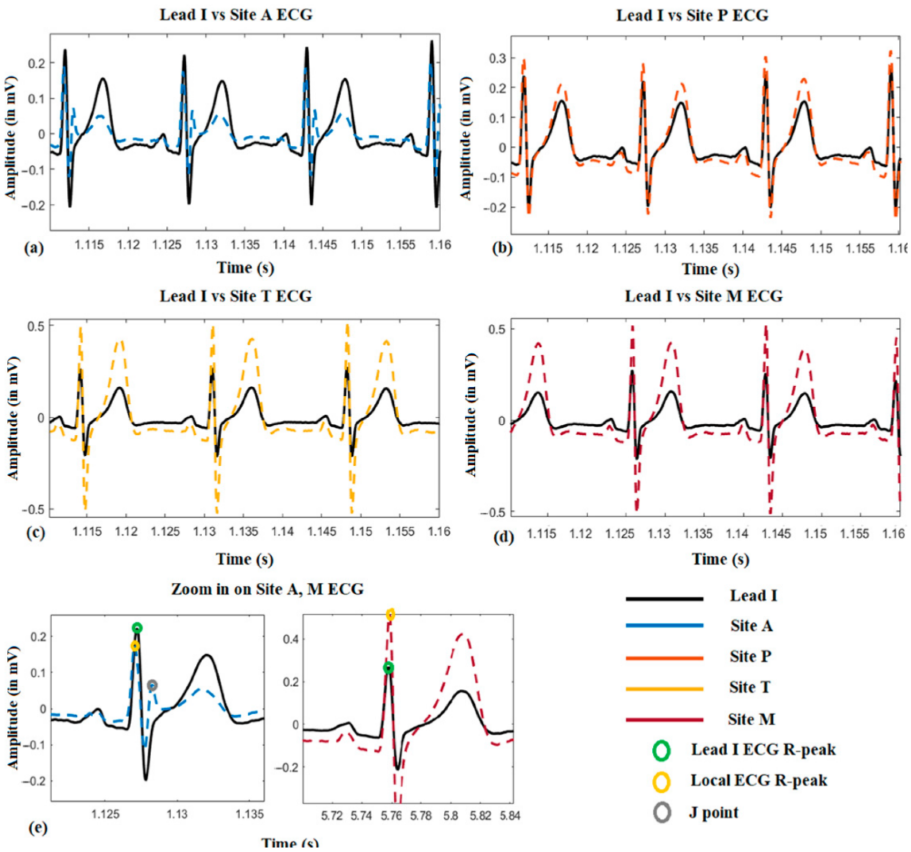


Fig. 7.3 The local ECG morphological variation compared with reference Lead I ECG.

2) Location Effect on EDR among the Auscultation Sites

Table 7.1 summarized the EDR MAE of each subject averaged across estimation techniques and given per auscultation site. ANOVA results indicate that there is no statistical difference between the five sites ($p = 0.746$), and there was no interaction between EDR algorithms and sites ($p = 0.516$). All four EDR algorithms have quite close MAE between each auscultation sites, including average MAE at A: 1.656 ± 0.351 bpm (breath per minute), P: 2.297 ± 0.476 bpm, T: 1.733 ± 0.461 bpm, M: 1.467 ± 0.326 bpm, and reference ECG (Lead I): 1.834 ± 0.378 bpm. This indicates that RR can be harvested using ECG anywhere on the chest with negligible location effect. Fig. 7.4 further visualized the location effect with different algorithms.

3) The Performance of the EDR Algorithms

After statistical analysis of the MAE on each subject with different EDR methods shown in Table 7.2, there was a significant difference between the four EDR algorithms ($p <$

Table 7.1 The ECG-derived respiration (EDR) mean absolute errors (MAE) of each subject for each auscultation site and Lead I ECG signals, averaged across estimation techniques in bpm.

	A	P	T	M	Lead I
Subject1	0.16	0.06	0.14	0.12	0.08
Subject2	2.27	3.74	1.77	0.86	2.00
Subject3	1.88	1.37	3.17	2.15	2.60
Subject4	0.50	1.69	0.54	0.82	0.83
Subject5	1.27	1.10	0.28	0.14	0.47
Subject6	5.79	4.61	6.91	3.23	5.45
Subject7	1.60	2.60	0.38	2.60	2.24
Subject8	2.54	5.41	3.16	3.47	2.24
Subject9	0.73	0.20	0.67	0.53	1.16
Subject10	0.36	1.62	1.69	1.25	1.11
Subject11	0.85	2.10	1.03	1.46	1.45
Subject12	1.93	3.06	1.07	0.99	2.39
Mean	1.66	2.30	1.73	1.47	1.83

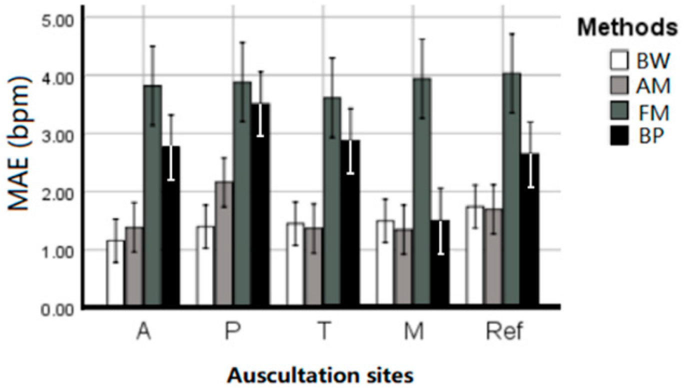


Fig. 7.4 The performance (MAE ± SE bpm) of the EDR algorithms on different auscultation sites and Lead 1 ECG signals.

0.001). The BW algorithm performed with MAE = 1.446 ± 0.181 bpm, closely followed by the AM algorithm with 1.589 ± 0.1966 bpm. Post hoc analysis revealed no statistical difference between BW and AM ($p = 0.31$), however, they were both significantly better ($p < 0.05$) than BP (MAE of 2.656 ± 0.258) and FM (MAE of 3.855 ± 0.329 bpm).

4) Time vs. Frequency Domain

Deriving respiration rate using the median frequency (overall MAE 1.80 ± 0.223 bpm) outperformed the counting method (overall MAE 2.98 ± 0.312 bpm) in the time domain ($p < 0.001$) suggesting stability of the frequency domain, although a significant interaction ($p < 0.001$) with the applied method was observed. From Fig. 7.5, it can be seen that

Table 7.2 The ECG-derived respiration mean absolute errors of each subject on different ECG-derived respiration methods, averaged across auscultation sites in bpm.

	BW	AM	FM	BP
Subject1	1.61	1.92	5.86	1.93
Subject2	1.65	2.88	3.26	2.45
Subject3	1.33	1.82	2.04	0.47
Subject4	2.19	1.99	4.06	2.19
Subject5	0.91	0.92	1.54	2.93
Subject6	0.49	0.41	3.55	3.67
Subject7	0.34	0.66	0.73	0.86
Subject8	0.76	0.75	3.23	1.19
Subject9	3.98	3.54	11.93	6.75
Subject10	0.45	0.38	2.08	0.38
Subject11	1.81	2.02	3.36	3.59
Subject12	1.84	1.78	4.61	5.89
Mean	1.45	1.59	3.85	2.69

the median frequency can provide a more accurate estimated RR on BW, AM, and FM algorithms. However, counting in the time domain is more accurate for the BP algorithm.

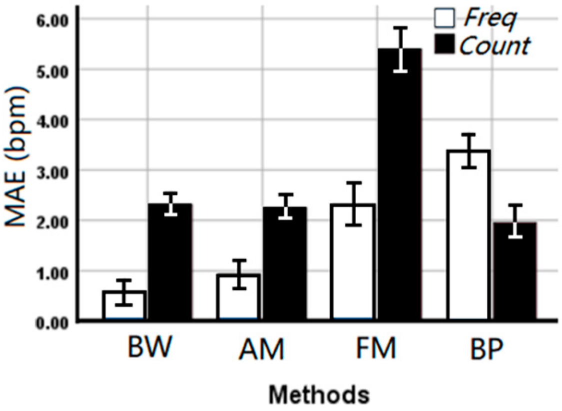


Fig. 7.5 The mean absolute error (MAE ± SE bpm) of the EDR rates between EDR algorithms and estimation methods. Freq: Median frequency in the frequency domain, Count: Automatic counting in the time domain.

7.2.4 Discussion

This study aimed at analysing the performance of one-lead EDR algorithms in auscultation site ECG signals and EDR rate estimation in both time and frequency domain. The results show:

Firstly, it is found that the location effect on the obtainment of EDR between auscultation site and Lead I ECG signals is negligible in our experimental data. The result revealed that the ECG morphological variation between auscultation sites happened on the onset and amplitude of the ECG components including the R-peak delayed from site A to M, and the amplitude increase of R-peak, s-wave, and T-wave. These won't directly affect the EDR signals extraction, but it is worth noticing in cardiac researches. It is still unclear on the occurrence of J-point elevation or RSR' (An ECG finding in which there are two R waves) in five subjects' site A ECG signals. Normally they are pathological, but the subjects were confirmed healthy with no heart conditions, and this can be normal for the age group. The high J-point or double R-peaks may interfere with R-peak detection when the fake R-peak is higher than the true one. In our study, the performance of the four chosen EDR techniques was not affected, however, it may have an impact on the QRS area or slope based EDR methods. In the study of Sakai, it indicated that the location of the electrodes affected the quality of EDR signals and the more accurate RR estimation was obtained when the electrodes were attached near the heart [157]. The best placement was a negative electrode at the bucket-handle and a positive electrode at pump-handle movements of the ribs. However, in our experiment, electrodes were placed at auscultation sites on the upper chest, which were already close to the heart. Besides, as we want to design a miniaturised device, the inter-electrode distance was fixed and short (10 cm). Therefore, from the physiological mechanism, the locations in our study barely have an effect on the respiratory sinus arrhythmia, and the effect on the respiration-induced electrical axis rotation and chest undulation-induced baseline wander are minimal. This result verifies that the location effect on RR estimation can be ignored in designing an integrated cardiac screening device.

Secondly, the BW and AM algorithms outperformed FM and BP algorithms. Although BW has a slightly smaller MAE (1.446 ± 0.181 bpm) than AM (1.589 ± 0.1966 bpm), the difference ($p = 0.315$) is not statistically significant in our experiment data, which cannot confirm that the performance of BW is better than AM so far. This is in contrast with some previous work. In Charlton's study, it was shown that the BW performed better than AM without statistical analysis [144]. The performance of FM and BP methods are in line with previously reported MAE using the PhysioNet's MIMIC-II database, while the results of AM obtained in this study are similar to the MAE reported by [159] using their experimental data. It seems like the experimental setting for the database had a significant impact on the performance of EDR algorithms. Because our experiments were conducted under ideal conditions where the subjects were required to lie down calmly without any movement, the MAE was much lower than studies that have made use of the database.

For respiratory sinus arrhythmia induced FM, the magnitude of the oscillation varies from individual to individual, so that the obtained EDR signal is not that conspicuous sometimes [158]. For example, the FM waveform of 0–50 s is shown in Fig. 7.2, the EDR signal in that period is messy, thus it can dramatically affect the peak detection in the time domain, causing inaccurate RR estimation. That should be the reason for FM's poor performance. For the BP algorithm, the choice of the frequency band is the current limitation. Though the frequency band (0.1–0.5 Hz) used in this study is appropriately extended, it is still not enough to capture RR from young children and stress tests. Besides, the use of a simple band-pass filter cannot remove unwanted interferences completely. The low-frequency component between 0.1–0.2 Hz, which is related to the baroreceptor reflex (blood pressure is regulated by the baroreceptors through the autonomic nervous system) and the high-frequency harmonic between 0.4–0.5 Hz interfere with the RR estimation in the frequency domain [255]. Therefore, an adaptive frequency band is essential to improve the performance of using a band-pass filter.

Thirdly, the RR estimation in the frequency domain is found to be better than the time domain for BW, AM, and FM [256]. This result is the opposite of Charlton's result, which said Fourier analysis was inferior to breath detection in the time domain [144]. As discussed above, there are conditions where the EDR is not conspicuous enough, thus in the time domain, it is hard to detect the corresponding respiration related peaks, while still possible to capture it based on the power spectral density function. Besides, at the beginning and end of the EDR signals, there may be incomplete breathing, this leads to the error for counting in the time domain. As there are not many breaths per minute, these errors are considerable for the RR estimation. Using frequency features reduces this error moderately. However, it is also noticed that the performance of frequency estimation for the band-pass filter is worse than counting in the time domain as there are mentioned lower-frequency and higher-frequency components in the spectrum which weaken the domination of the respiratory band. Therefore, further analysis of frequency components is needed to improve accuracy when using the BP algorithm.

In this study, our research focused on the EDR of healthy adults at rest, and the experimental conditions were ideal that the subjects kept supine and breathed evenly without any movement. However, there are conditions of practical application that need to be considered including EDR performance on irregular respiration, such as deep breath or an increased respiration rate. Improvement still can be done to improve the RR estimation accuracy. The current validation study has compared four algorithms of the existing algorithms. More algorithms and fusion methods could be tested to improve the accuracy for clinical use. Future studies can include RR estimation throughout monitoring via a Holter-like monitor.

This study analysed the location effect on EDR algorithms' performance between auscultation sites and compared four EDR algorithms to estimate RRs in the time and frequency domain. The results showed that, firstly, the location of the ECG electrodes between auscultation sites barely affects the estimation of RR. Secondly, the BW and AM algorithms outperformed than FM and BP algorithms in generating the approximation of the respiratory signal. Thirdly, RR estimation in the frequency domain is more reliable except on BP algorithms. All the findings contribute to building chest-based multiple physiological parameter monitors and providing more accurate RR estimation using EDR algorithms.

7.3 Comparison between Embroidered and Gel Electrodes on ECG-Derived Respiration Rate

7.3.1 Introduction

In the design of wearable ECG devices, the choice of sensor is very important. The commonly used gel electrode is not suitable to wear for a long time, because it can cause skin discomfort. There are a number of benefits to using an embroidered electrode instead of a gel electrode: 1) comfort in wearing, 2) reusability, 3) unobtrusiveness when sewed on the clothes, 4) low-cost [257]. In recent years, textile electrodes are gaining application in different fields, such as capturing the electromyography (EMG) [257–259], electroencephalography (EEG) [260, 11], and electrocardiogram (ECG) [10], [11]. There are also studies proposed to extract respiratory signals from ECG using textile electrodes [261, 262]. However, their performance in capturing ECG-derived respiration (EDR) signals are not systematically compared with gel electrodes. Besides, embroidered electrodes also have the drawback, as the captured signal quality is usually affected by the fit to the skin, so the signals collected by embroidered electrodes are usually not as good as the gel electrodes [263]. Therefore, in this study, we analyze the effect of using embroidered electrodes to obtain the EDR signals.

The primary aim of this study was to compare the performances of embroidered electrodes and gel electrodes on deriving EDR signals. The second aim was to investigate the effect of using mean frequency, median frequency, peak frequency and counting in the time domain to estimate RR. All the findings help consolidate the use of embroidered electrodes in wearable medical devices and provide the basis for designing more accurate EDR algorithms.

7.3.2 Methodology

1) Subjects

The experiments were conducted on nine healthy human subjects with no history of heart diseases (6 male/ 3 female, age range 21–28 years, mean 25.2 years). The procedures were approved by the King's College Research Ethics Committee (Approval No.: LRS-18/19-10673). Subjects gave written informed consent prior to the experimental procedures.

2) Experimental Setup

The textile electrodes used in the experiment were developed at the Centre for Robotics Research (CORE) at Kings College London. They are made of silver coated thread (Electro-Fashion Conductive Thread, Kitronik, $40\Omega m^{-1}$) embroidered into linen fabric with a Vilene cut away stabiliser. The embroidered pattern used is a 20 mm diameter circle with a cross hatched fill pattern. The hatching has a 2 mm separation and two iterations of embroidery are performed [257]. The solid gel electrodes used in the experiment were Ambu WhiteSensor WS (36 × 40 in mm). They were all positioned over the chest of a subject with standard Lead I for concurrent acquisitions of the ECG signals. The embroidered electrodes were fixed with medical tapes to enhance the adhesion. A microphone was placed under the subjects' nose to record the actual respiration as a reference. The recording used the commercial acquisition system (iWorx, model RA 834) and ECG devices (iWire-BIO4) as recorder. The sampling frequency was 1 kHz and the filter for the ECG was 0.05 – 40 Hz. A simple block diagram of this hardware system is shown in Fig. 7.6.

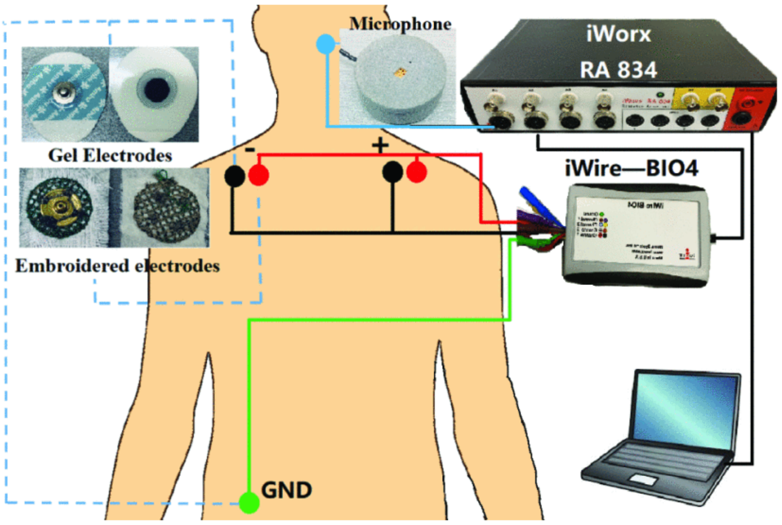


Fig. 7.6 Block Diagram of the recording setup: Red dots (textile electrodes), black dots (gel electrodes), the green dot is GND using gel electrodes. The order of red and black positions was randomly assigned. The blue dot is the microphone for respiration recording. iWire BIO4 is for ECG recording. All the data is transferred to the computer for processing through iWorx RA 834.

During the experiments, all the subjects should keep supine and remain calm. The duration of the recording was 3 minutes, repeated 3 times, totaling 9 minutes worth of data. Breaks of 2 minutes were given between trials.

3) Respiration signal extraction method

In this study, the baseline wander of the ECG was used to capture the EDR signals, which has shown to perform better than the other methods [144]. The processing was conducted in the Matlab® R2018b environment and the statistical analysis was performed using IBM® SPSS version 26.

The captured ECG signals and respiration sound were filtered first to remove the unwanted noise. For the ECG, A zero-phase 3^{rd}-order Butterworth high-pass filter with 0.1 Hz was used to eliminate the artifacts. For the respiration sound, a 3^{rd}-order Butterworth band-pass filtered (0.1 - 0.5 Hz) was used to smooth the waveform.

The Pan–Tompkins algorithm was used to capture the R peaks in the ECG signals, then found the Q points using the gradient descent method based on R peaks [89]. The baseline wander can be generated by connecting the middle points of the R peaks and Q points [159]. A representative derived respiration signals (baseline wander) and reference respiration signal are shown in Fig. 7.7(b).

The numbers of breaths were counted in the recorded respiration sound manually to ensure the accuracy, then the reference RRs were obtained by the counted numbers divided by the experiment duration (3 mins). The EDR rates were calculated using the counting method in the time domain, and frequency features. For the counting method, the respiration signal was interpolated to the same sample size of its raw ECG signal first to increase the signal resolution. Then, we used a moving average filter (window length: 50 ms) to smooth the respiration signal and automatically detected the number of peaks with the threshold of the signal mean value to get the RR of the ECG duration. For the frequency methods, the mean frequency, median frequency and peak frequency of the EDR signal between 0.1 to 0.5 Hz were calculated and multiplied by the time duration to get RRs. The spectrum of the representative EDR signals are shown in Fig. 7.7(c).

The mean absolute errors (MAE) between EDR rates and reference RRs were calculated. Then, using a two-way repeated measures analysis of variance (ANOVA) with factors electrode materials and RR methods, performances could be compared.

7.3.3 Results

Results for the embroidered and gel electrodes EDRs were presented in Fig. 7.8. The P value between electrode materials was 0.077, which means there was no statistical difference between the performance of embroidered electrodes and gel electrodes in capturing EDR signals. However, the absolute values of the errors were lower for gel electrodes compared to embroidered electrodes. For the difference between the estimation methods, the P value was 0.01, which means there was a statistical difference between the methods. From Fig. 7.8(b), the median frequency provides the smallest MAE among

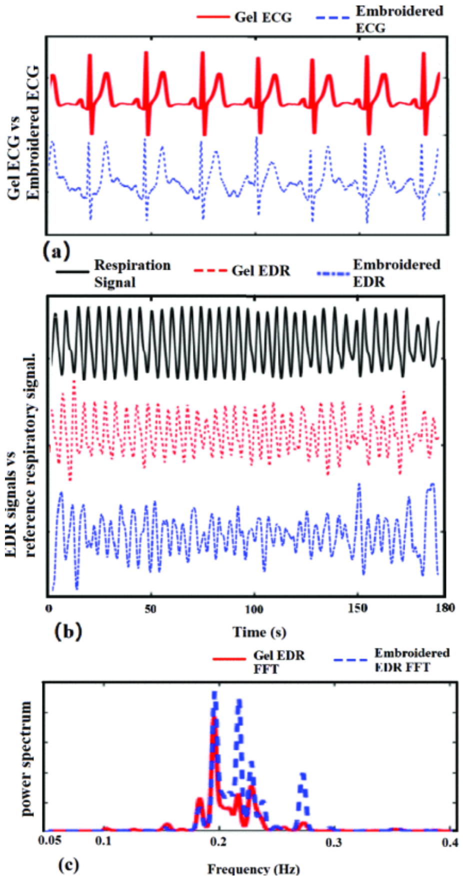


Fig. 7.7 (a) A segment of the ECG signals captured by embroidered and gel electrodes; (b) A representative derived respiration signals (baseline wander) and reference respiration signal; (c) The spectrum of the representative EDR signals in (b).

the methods. No interaction was found between electrodes and methods and post hoc analysis revealed no difference among frequency features, but both the mean and median frequencies performed better than the counting method.

7.3.4 Discussion

1) Embroidered electrodes with gel electrodes

The aims of this study was to investigate the performance of embroidered electrodes in capturing EDRs and analyse which frequency feature best estimates RR. From the results, we can see that the embroidered electrodes can be used to capture EDR signals, because there is no statistical difference compared with gel electrodes. However, from the MAE and standard error (SE), it shows the embroidered electrodes performance not as good as Gel electrodes in capturing EDRs. By analyzing specific data, we found that the cause of the poorer performance of embroidered electrode EDR was the failure of R-peak detection. Because the embroidered electrodes require a close fit with the body,

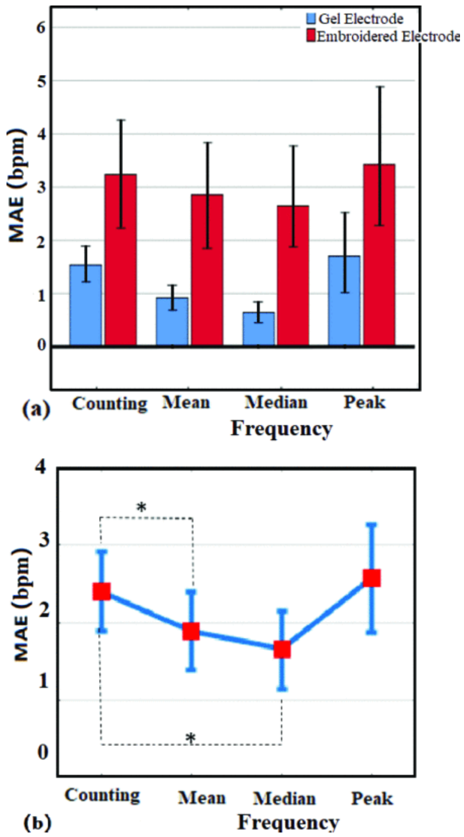


Fig. 7.8 (a) The mean absolute error (MAE ± standard error, SE, measured in breaths per minute, bpm) of the EDRs between methods and electrode types. (b) The interaction between the RR estimation methods (MAE ± SE). The difference is significant at the 0.05 level is marked with *.

huge artifacts were generated when the contact was poor. Some artifacts were not filtered during pre-processing, which caused the Pan-Tompkins algorithm (Squaring) to fail. These artifacts can be eliminated by further digital signal processing, enhancing the performance of the embroidered electrode EDR.

We also calculated the p value between the two types of electrodes after removing the 4 large error group data (caused by R peak detection failure), which increased from 0.077 to 0.666. That means the performance between the two types of electrodes on EDR would be closer. Because our experiments were conducted under ideal conditions that the subjects were required to lie down calmly without any movement, the MAE of gel electrodes (1.21 ± 0.72 SE bpm) in our results was smaller than the previous studies (4.87 in counting/ 7.51 in Fourier bpm) [144]. But the performance of embroidered electrodes in obtaining EDR (MAE 3.05 ± 1.83 SE bpm) was not good compared with the results of Park et al. (MAE 1.13/ 1.09 bpm for two types of textile electrodes) and Shen et al. (MAE 1.4 - 2.3 bpm) [264, 262]. Although the comparison cannot directly reflect the performance of our embroidered electrodes due to the different experimental environments. We can still find that the size of the embroidered electrode and appropriate placement can enhance their stability. In the future study, the stress testing, such as walking and running can also be conducted to examine the performance of embroidered electrodes in capturing EDR.

2) Frequency feature on EDR

According to our results, the choice of method has an impact on the estimated RRs. From Fig. 7.8, we can also see that using the median frequency to calculate EDR is more accurate than the other methods. This result is different from that proposed in Charlton's paper [144]. His results show that counting in the time domain is more accurate than Fourier in the frequency domain because the low-quality segments of the signals were eliminated before processing. However, our results prove that the EDR estimation in the frequency domain can work even better. Besides, frequency estimation is more robust, especially when the electrodes are in poor contact for a short period, and small segment of the ECG signal is lost. This barely affects the EDR in the frequency domain, but for counting in the time domain, it may lose the number of breaths during that period. In the frequency domain, the peak frequency method did not perform well is due to the RR is mutative, which can be represented as multiple peaks in the spectrum. Therefore, only using the highest peak can cause an error on the estimated RRs, and the error can become larger when breath irregularly. For the mean and median frequency, our results show that the median frequency provides more accurate RRs.

The study found that using embroidered electrodes can provide highly accurate RRs as gel electrodes, but proper electrode-skin contact is the prerequisite to ensure EDR performance. Besides, we found that using median frequency of the obtained EDR signals outperformed the counting method. All the findings contribute in building medical device using embroidered electrodes and providing more accurate RR estimation using EDR signals in the frequency domain.

7.4 Verification of the ECG-Derived Respiration Rate Performance under Stress Test

7.4.1 Introduction

As discussed in Section 2.4, most of the current EDR researches are mainly aimed at the static, and few analyze under stress test. However, cardiac stress test can reflect the abnormal coronary blood flow in heart muscle tissue when the subjects are maximally exercising. The results have great significance for evaluating some hidden or early cardiac diseases, such as coronary artery disease. Besides, the stable and reliable EDR methods under stress test can also enhance the monitoring and portability of the future ECG-

PCG integrated devices by providing real-time respiratory rate. The preliminary research completed so far focused on the traditional 12-lead ECG. The specific research aims include: (1) To verify the EDR performance under stress test by comparing with static; (2) to investigate whether the selection of ECG lead has an impact on the extraction of EDR; (3) to compare the commonly used EDR methods and find out a more suitable method for stress test. All the findings contribute to verifying the feasibility of EDR technology under stress test, and designing a more accurate and trustworthy EDR methods.

7.4.2 Methodology

1) Subjects

The experiments were performed on 7 healthy human subjects (5 male/2 female, age range 22-24 years) with no history of heart diseases or respiratory issues. The procedures were approved by the King's College Research Ethics Committee (Approval No.: LRS-18/19-10673). Subjects gave written informed consent before the experimental procedures.

2) Experimental Setup

There were two phases during the experiments. In Phase One (static test), subjects kept supine. Their 12-leads ECG signals and oronasal pressure breathing signals were recorded simultaneously. The 12-leads ECG include six precordial leads (V1, V2, V3, V4, V5 and V6), three standard limb leads (I, II and III), and the augmented limb leads (aVR, aVL and aVF). The sensors for ECG signals were solid gel electrodes (Ambu WS, size: 36 × 40 mm, Medico Electrodes International LTD., Uttar Pradesh, India), and the roronasal pressure breathing signal was captured by oxygen mask with hose. The recording used the commercial acquisition system (iWorx, model RA 834, iWorx Systems Inc, Dover, NH, US) and ECG devices (iWire-ECG12, iWorx Systems Inc, Dover, New Hampshire, US). The recording lasted 10 minutes.

In Phase Two (stress test), subjects were required to ride a fitness bike at a constant rate. In the meantime, their ECG signals and breathing signal were also recorded for at least 8 minutes. When the heart rate had exceeded 50% than static, the acquisition can be early-stopped to guarantee the health of participants.

Besides, to reduce the noise effect, silence were kept during the whole experiments. The sampling frequency of the recording device was 200 Hz. Fig. 7.9 shows the experimental settings during each phase.

3) Signal Processing

In this study, we also used BW, AM and FM and as EDR features. The signal processing method was the same as that in 7.2.2, so it is not be repeated here. Considering that there were huge artefacts in the ECG signals under stress test, in this study, we applied a moving window to extract EDR to reduce the artefact effect. The window size was 1 minute, and

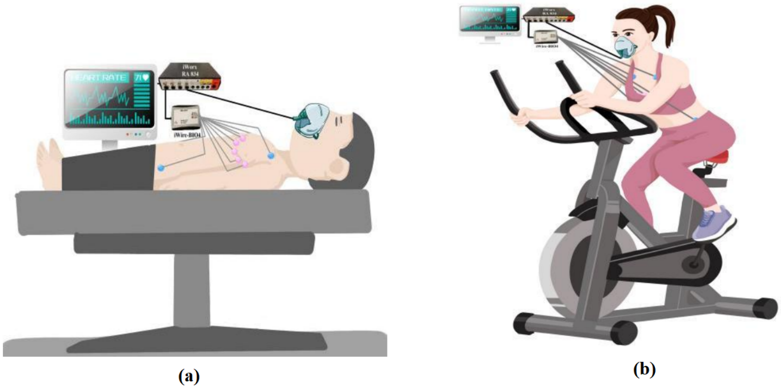


Fig. 7.9 Experimental settings for each phase. (a) static test; (b) stress test.

the moving step was 20 seconds. In the result analysis, the averaged EDR rate of the experimental signals was used. MAE between EDR rate and reference respiratory rate was calculated to evaluate the performance.

7.4.3 Results

1) Stress test vs. Static

Table 7.3 summarizes the EDR performance under static test and Table 7.4 shows the performance under stress test. For static test, the overall MAE is approximately 2.17 bpm. In Table 7.4, the '-' means the EDR was failed due to unsuccessful R-peak detection. In the MAE calculation, the EDR rate was taken as 0. The overall MAE is approximately 5.48 bpm which is dramatically larger than static test and cannot be regarded a trustworthy respiratory rate.

2) ECG Lead Selection Effect on EDR

Fig. 7.10 shows the EDR performance with different ECG leads under static test and stress test. Under static test, the EDR at precordial (V1 - V4) leads performed better. Under stress test, considering the huge errors group by failed EDR extraction, it cannot intuitively show the difference. However, lead I and aVL performed worse due to failure in almost all R-peak detection.

3) Comparison of the EDR Methods

After calculation from Table 7.3 and Table 7.4, it is found that under static test, BW (averaged MAE 1.40 bpm) performed better than AM (averaged MAE 2.05 bpm) and FM (averaged MAE 3.03 bpm). However, under stress test, the three methods performed all bad, with averaged MAE 5.65 bpm, 5.82 bpm and 4.95 bpm, respectively. The difference between them are also not statistically significant.

Table 7.3 EDR Performance (MAE) under static test in bpm.
S: Subject, Res: Reference Respiratory Rate

	I	II	III	V1	V2	V3	V4	V5	V6	aVR	aVL	aVF	Res
S1—BW	0.12	4.52	0.06	0.34	0.30	0.30	0.58	0.71	0.06	0.32	0.09	1.34	17.80
S1—AM	0.05	0.48	0.10	0.39	0.83	0.69	0.78	0.63	0.19	0.47	0.18	0.25	
S1—FM	0.41	0.41	0.45	0.35	0.39	0.42	0.43	0.41	0.37	0.38	0.39	0.34	
S2—BW	0.18	1.19	0.26	0.24	0.30	0.34	0.32	0.27	4.61	0.34	1.21	2.37	17.20
S2—AM	1.29	5.98	4.20	0.22	0.29	0.32	0.87	0.59	6.23	7.01	6.59	6.14	
S2—FM	0.26	0.27	0.27	0.26	0.26	0.27	0.27	0.27	0.27	0.26	0.37	0.27	
S3—BW	1.04	3.26	0.52	0.55	0.58	0.67	0.67	1.92	6.62	3.01	1.36	7.58	17.43
S3—AM	0.69	7.77	1.71	3.46	0.85	0.79	1.10	2.35	7.60	7.48	7.29	7.71	
S3—FM	6.02	5.74	5.10	5.46	5.53	5.71	5.67	5.67	5.21	6.05	5.46	5.45	
S4—BW	5.47	1.39	2.33	0.82	1.96	2.11	2.10	3.14	1.29	1.30	1.83	1.31	19.19
S4—AM	5.53	1.34	1.83	1.00	1.41	1.37	1.51	1.85	1.35	1.26	1.35	1.34	
S4—FM	8.03	7.31	7.02	6.86	7.28	7.59	7.34	7.35	7.23	7.98	7.77	8.19	
S5—BW	1.63	1.35	1.39	1.23	1.32	1.60	1.35	1.37	1.43	1.07	1.68	1.40	13.84
S5—AM	1.67	1.33	1.20	0.46	1.15	1.29	1.44	1.44	1.37	1.26	1.11	1.35	
S5—FM	1.98	1.93	1.91	1.90	1.92	1.96	1.92	1.92	1.92	1.11	1.92	1.96	
S6—BW	1.31	3.92	0.97	0.70	1.27	1.03	1.28	2.14	2.31	0.31	1.43	2.97	17.45
S6—AM	1.40	2.82	0.75	0.83	1.23	1.01	1.11	1.34	1.85	1.02	1.21	2.41	
S6—FM	5.91	5.33	4.79	5.94	5.70	5.52	5.50	5.32	5.28	5.04	5.64	5.29	
S7—BW	0.23	0.79	0.28	0.45	0.32	1.36	0.22	0.43	1.90	0.28	0.17	1.81	14.52
S7—AM	3.74	2.51	4.24	0.32	0.24	2.80	1.15	0.38	2.48	2.85	2.64	2.32	
S7—FM	0.21	0.21	0.20	0.21	0.21	0.21	0.21	0.22	0.21	0.22	0.22	0.21	
Mean	2.25	2.85	1.88	1.52	1.59	1.78	1.71	1.89	2.85	2.34	2.38	2.95	

Table 7.4 EDR Performance (MAE) under stress test.
S: Subject, Res: Reference Respiratory Rate, -:Failed EDR extraction by unsuccessful R-peak detection.

	I	II	III	V1	V2	V3	V4	V5	V6	aVR	aVL	aVF	Res
S1—BW	0.02	1.12	0.89	0.03	2.45	0.46	0.04	0.57	6.23	0.66	7.89	2.14	
S1—AM	0.12	0.53	3.35	1.49	3.31	2.39	1.19	0.29	6.02	0.66	8.46	2.14	16.26
S1—FM	0.64	1.32	1.20	0.91	0.41	0.50	0.52	0.94	5.44	1.64	8.42	0.68	
S2—BW	-	1.31	1.93	0.84	2.52	1.17	1.69	0.81	0.73	-	-	0.99	
S2—AM	-	1.86	2.37	0.47	1.82	1.77	0.64	0.54	1.30	-	-	1.09	12.57
S2—FM	-	1.38	0.16	0.32	0.59	0.84	0.29	0.30	0.28	-	-	0.29	
S3—BW	1.48	0.12	0.81	1.72	0.27	0.99	0.86	1.64	0.25	1.11	0.51	0.17	
S3—AM	2.43	0.13	0.74	1.83	0.64	1.25	1.04	1.85	0.12	1.54	0.87	0.01	11.86
S3—FM	0.28	1.18	1.25	1.26	1.32	1.35	1.37	1.34	0.78	1.24	2.12	0.02	
S4—BW	-	8.82	7.56	7.93	7.74	11.47	11.71	10.94	10.24	10.93	-	9.79	
S4—AM	-	9.97	7.11	9.65	7.08	10.45	11.45	11.03	10.20	11.11	-	11.92	22.16
S4—FM	-	7.52	7.37	14.51	6.41	14.53	13.74	12.08	9.54	7.77	-	6.65	
S5—BW	-	3.69	2.65	2.38	4.54	0.15	5.15	4.50	4.43	17.96	3.54	4.23	
S5—AM	-	3.82	2.50	0.07	2.65	1.93	5.25	5.47	4.45	17.96	2.21	4.50	17.95
S5—FM	-	5.35	2.95	5.86	4.96	9.69	7.95	6.48	8.92	17.96	3.75	8.70	
S6—BW	-	9.40	9.52	6.55	10.08	8.58	8.15	8.58	8.27	9.46	-	9.30	
S6—AM	-	9.49	9.67	8.82	10.09	8.55	8.14	8.60	8.48	9.64	-	9.30	18.13
S6—FM	-	0.29	0.36	1.68	0.29	1.84	0.30	0.29	2.55	2.07	-	1.50	
S7—BW	-	0.77	-	3.08	1.83	2.78	2.18	2.09	1.00	2.40	-	0.23	
S7—AM	-	0.93	-	2.99	2.37	2.83	2.60	1.98	1.18	2.65	-	0.19	13.17
S7—FM	-	0.45	-	0.73	0.01	0.45	0.45	0.45	0.13	2.23	-	0.44	
Mean	12.24	3.31	4.85	3.48	3.40	4.00	4.03	3.85	4.31	-	11.23	3.54	

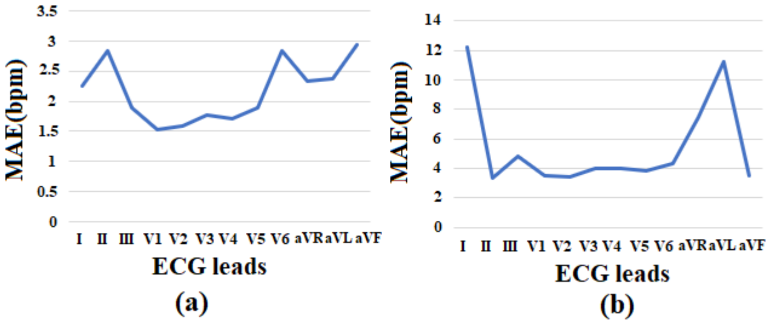


Fig. 7.10 EDR performance with different ECG leads. (a) under static test; (b) under stress test.

7.4.4 Discussion

For EDR under static test, the overall MAE 2.17 bpm basically stands in line with the research in Section 7.2. Besides, after specific analysis, it can also be found that the performance of the overall EDR under the static test is relatively stable and reliable except for a few results with large errors. However, under stress test, the overall MAE 5.48 bpm is not satisfying and cannot be regarded as trustworthy respiratory rate. Though, medical taps were used to fix the ECG wire reducing the swing, the artefacts were still large and cannot be easily remove by processing techniques. As a result, the huge movement artefact caused R-peak detection failure leading failed EDR extraction. However, this does not mean that EDR is completely unfeasible under stress test. Because it worth noticing that on Subject 1, 2 and 3, the EDR performed satisfying with tiny error on specific leads. Therefore, we can basically conclude that reducing artefacts by wire swing is critical for applying EDR under stress test. Using wireless ECG may be a potential surrogate.

In static test, the precordial leads (V1 - V4) seemed to have better performances compared with limb leads. Because they may perceive the breathing induced chest undulation better, and affected less by the small unconscious movements of the body. However, in stress test, it is unlikely to say which lead performed better so far, but Lead I and aVL obviously performed worse. After analyzing the waveforms of the 12-lead ECG signals, it is found that the amplitude of Lead I ECG is usually smaller. As a result, it is easier to be obliterated when there is huge artefact. aVL is derived from Lead I, so the same problem may exist.

In the comparison of the EDR methods, under static test, this study got the same result as Section 7.2 that BW performed better than AM and FM. However, under stress test, since these methods all relied on R-peak extraction, it was unlikely to evaluate their pros and cons so far. More EDR methods, such as EMD, can be tried in follow-up research.

In addition to the above findings about EDR, we also found that comparing static and stress tests, only one subject had a significant increase in respiratory rate (from 19.19 to 22.17 bpm), while the rest basically remained or even decreased. The analysis of the breathing signal revealed that people actively adjust to deep breathing to maintain the respiratory rate during exercise, and only increase the rate when it cannot be maintained. It is reflected in the respiration signal as the amplitude of the signal becomes larger. This knowledge can be also helpful in the design of future experiment protocol.

7.5 Conclusion

This chapter discusses the works conducted on EDR techniques for eliminating the necessary of extra respiratory sensor. In Section 7.2, we experimentally analyze whether the accuracy of ECG derived RR depends on the auscultation sites. Experiments were conducted on 12 healthy subjects to obtain simultaneous ECG (at auscultation sites and Lead I as reference) and respiration signals from a microphone close to the nostril. Four EDR algorithms were tested on the data to estimate RR in both the time and frequency domain. Results reveal that: (1) The location of the ECG electrodes between auscultation sites does not impact the estimation of RR, (2) baseline wander and amplitude modulation algorithms outperformed the frequency modulation and band-pass filter algorithms, (3) using frequency domain features to estimate RR can provide more accurate RR except when using the band-pass filter algorithm. These results pave the way for ECG-based RR estimation in miniaturised integrated cardiac screening device.

In Section 7.3, we compared embroidered electrodes with gel electrodes on their performance in capturing EDR signals and analysed which frequency feature best estimates RR. Data were collected from 9 healthy subjects. Results reveal that (1) embroidered electrodes performed similarly to gel electrodes (P = 0.077), (2) using the median frequency of the obtained EDR signals is significantly better (P = 0.01) than the counting methods in the time domain. The obtained results are relevant for the future development of textile-based sensors.

In Section 7.4, we discussed the feasibility of EDR under stress by experiments. Data were collected from 7 healthy subjects. Current results revealed that the wired ECG is not reliable for EDR under stress test with huge artefacts. In the future study, we may test wireless ECG and more EDR methods.

Chapter 8

Conclusion and Future Work

8.1 Conclusion

This thesis aims at advancing the design of a multi-sensor miniaturised device for cardiac investigation and monitoring. The main works focus on : (1) developing machine learning techniques to improve the computer-aided diagnosis of PCG and ECG; (2) assessing the feasibility by experiments, utilising biomedical signal processing to eliminate the need of physical respiration sensor. Specifically, the contributions of this thesis include:

1) Assess the feasibility of localised ECG signal acquisition and analyse its usability for PCG segmentation. In Chapter 4, we experimentally investigated the time property of ECG and PCG signals at auscultation sites and the effect of ECG inter-electrode distance. Results showed that ECG signal could be acquired stably at auscultation sites within a small area (5 cm), which provides a theoretical basis for designing miniaturised integrated ECG-PCG devices. The inter-distance of the electrodes for ECG does not affect the occurrence time of the R-peak. Besides, the delay between R-peak and onset of first heart sound (S1) depends on the auscultation site e.g. S1 onset occurs before the R-peak at auscultation site M. This study suggests that small integrated ECG-PCG devices can be made by reducing the distance between the ECG electrodes. In the meantime, distinguishing the auscultation location is necessary for performing more precise PCG segmentation using ECG as reference.

2) Investigate the optimal use of deep learning input and propose a reliable algorithm for PCG classification. In Section 5.3, a study was conducted aiming at analysing the duration effect on the commonly used deep learning methods to provide insight for future studies in data processing, classifier, and feature selection. The results of this study revealed that: 1. very short heart sound signal duration (1 s) weakens the performance of RNNs, whereas no apparent decrease in the tested CNN model was found. 2. RNN outperformed CNN using Mel-frequency cepstrum coefficients (MFCCs) as features. There was no difference between RNN models (LSTM, BiLSTM, GRU, or BiGRU). 3. Adding dynamic information ($\Delta \& \Delta^2 MFCCs$)) of the heart sound as a feature did not improve the RNNs' performance, and the improvement on CNN was also minimal ($\leq 2.5\%$ in MAcc). The findings provided a theoretical basis for further heart sound classification using deep learning techniques when selecting the input length. In Section 5.4, a comparative study was performed on deep convolutional neural networks (CNNs) to assess optimum time-frequency representations (TFRs) as input features for PCG classification. The results showed that the transformation of the heart sound signal into the TF domain achieves higher classification performance than using of raw signals. Among the TFDs, the difference in the performance was slight for all the CNN models (within 1.3% in average accuracy). However, Continuous wavelet transform (CWT) and Chirplet transform (CT) outperformed the rest. Besides, the appropriate increase of the CNN capacity and architecture optimisation can improve the performance, while the network architecture should not be overly complicated. Based on the ResNet or SEResNet family results, the increase in the number of parameters and the depth of the structure do not improve the performance apparently. In additional, combining TFDs as CNN inputs did not significantly improve the classification results. The findings of this study provided the knowledge for selecting TFDs as CNN input and designing CNN architecture for heart sound classification. In Section 5.5, a HMS-Net was proposed and won the first prize in the CinC/PhysioNet 2022 PCG classification challenge. Its recording quality assessment method was based on frequency density distribution for label correction to prevent the poor-quality recording segments from misleading network optimisation, and its network builds long short-term independencies between multi-scale features and improves the classification performance.

3) Propose a deep learning algorithm for the detection of paroxysmal atrial fibrillation using single-lead ECG. In Chapter 6, a two-stage RNN network was proposed during the China Physiological Signal Challenge 2021 (CPSC 2021) which had satisfying performance (90.14% and 92.56% accuracy on non-AFib and AFib segments classification respectively in the two stages) and held the advantage of small size (approximately 500 k). It showed promising potential for terminal equipment such as the miniaturized ECG-PCG device, personal laptop, mobile phone or embedded device. This will add side application for the future device.

4) Explore the feasibility to provide accurate respiratory rate while eliminating the need of physical respiration sensor. In Section 7.2, a study was conducted in the ECG-derived respiration (EDR) field to assess the feasibility of extracting EDR from the localised ECG at the auscultation sites by experiments. Results indicated that: 1. the location of the ECG electrodes between auscultation sites does not impact the estimation of RR. 2. baseline wander and amplitude modulation algorithms outperformed the frequency modulation and band-pass filter algorithms, and 3. using frequency domain features to estimate RR can provide more accurate RR except when using the band-pass filter algorithm. These results pave the way for ECG-based RR estimation in miniaturised integrated cardiac screening device. In Section 7.3, a comparative study was performed to test using embroidered electrodes to capture ECG signals for EDR extraction. Results reveal that: 1. embroidered electrodes performed similarly to gel electrodes (P = 0.077); 2. using the median frequency of the obtained EDR signals is significantly better (P = 0.01) than the counting methods in the time domain. The obtained results showed potentials in future development of low-cost textile-based sensors applications. In Section 7.4, we experimentally test the EDR performance under stress test. Results showed that the wired ECG is not reliable for EDR under stress test with huge artefacts, using wireless ECG may be a potential surrogate to improve the performance.

In general, this thesis verified the feasibility of ECG-PCG integrated device through experiments, and analyzed the segmentation of PCG using localized ECG as a reference. Using the open-source databases, exploration was conducted on the optimization of deep learning input to improve the PCG classification accuracy. Furthermore, with the prior knowledge in the exploration, a novel algorithm (HMS-Net) with recording quality assessment method was proposed which achieved a leading score in the PhysioNet 2022 Challenge. On ECG classification, a deep learning algorithm was proposed for atrial fibrillation detection, adding side application for the future device. To provide accurate respiratory rate while eliminating the need of physical respiration sensor, experiments were conducted to prove the possibility of providing reliable respiratory rate from the ECG-PCG device without adding an extra sensor.

8.2 Future Work

Although this thesis fully demonstrates the feasibility of miniaturized ECG-PCG integrated device, including sensor fusion, automatic auscultation and providing more physiological indicators. There is still work to do before deployment. Some potential future research directions are listed as follows:

1) To investigate the refined classification of PCG using machine learning techniques. So far, the vast majority of PCG classification researches still stay in the binary classification of whether there is murmur or not. However, the specific types of murmurs, the location of the murmurs, and the evaluation of the severity of the murmurs are still relatively blank. Furthermore, the existing databases are general for murmurs. The author thinks that more specific scenario, such as valve surgery tracking, congenital heart disease exacerbation with refined classification of PCG will greatly enhance the practicality of this technique.

2) To explore more diagnostic capabilities of single lead ECG signal using machine learning techniques. In this thesis, we explored using deep learning to do ECG classification on atrial fibrillation. Actually, there are more heart conditions can be detected by single lead ECG, such as supraventricular premature beat, supraventricular tachycardia, ectopic rhythm and atrial flutter. Therefore, there is still distance on the multi-class ECG classification. Besides, there are still remaining problems to be addressed from practical use in the automatic classification. Because the morphological and temporal characteristics of the ECG signals are significantly different from patient to patient and varies by physical conditions[71–73]. For instance, sometimes the morphologies are different for the same disease between patients or two different diseases show the same characteristics on ECG signals. In addition, for localized ECG, its morphology is likely to be different from standard ECG leads, so explore using localized ECG for classification is also a potential question to solve.

3) To conduct study of multi-channel PCG signals for classification improvement. In the case of heart disease, there should be interactions among the auscultation sites [265]. Theoretically, multi-channel simultaneous PCG measurement can reflect more comprehensive information of heart sound propagation in the cardiac region. For binary classification, this will contribute in identifying the existence of murmurs. In addition, multi-location measurements will also help determine specific murmur types and severity. In the PhyioNet 2022 database, although multi-channel data are also presented, they are not measured simultaneously. So, it is not conducive to building a model of murmur propagation. Therefore, it is worthy of continuing study in this aspect.

4) To improve the EDR performance under stress test In our experiments, EDR was not reliable in stress test due to the motion-induced artefacts. And, the artefacts are mainly caused by wobble on the ECG wire. Therefore, follow-up research can be carried out to test wireless ECG, or design a more novel method to complete EDR under stress test. In this way, ECG-based heart rate and respiration rate monitoring will greatly improve the daily monitoring ability and portability of the device.

References

[1] H. Wang, M. Naghavi, C. Allen, R. M. Barber, Z. A. Bhutta, A. Carter, D. C. Casey, F. J. Charlson, A. Z. Chen, M. M. Coates, *et al.*, "Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the global burden of disease study 2015," *The lancet*, vol. 388, no. 10053, pp. 1459–1544, 2016.

[2] World Health Organization, "Cardiovascular diseases (CVDs)." https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds), 2021.

[3] A. C. of Cardiology *et al.*, "Cardiovascular disease causes one-third of deaths worldwide:'alarming trends' for countries in all stages of development," *ScienceDaily. www. sciencedaily. com/releases/2017/05/170517143625. htm*, 2020.

[4] S. A. Lear, W. Hu, S. Rangarajan, D. Gasevic, D. Leong, R. Iqbal, A. Casanova, S. Swaminathan, R. M. Anjana, R. Kumar, *et al.*, "The effect of physical activity on mortality and cardiovascular disease in 130 000 people from 17 high-income, middle-income, and low-income countries: the pure study," *The Lancet*, vol. 390, no. 10113, pp. 2643–2654, 2017.

[5] W. H. Organization *et al.*, "Hearts: technical package for cardiovascular disease management in primary health care," 2020.

[6] E. Wilkins, L. Wilson, K. Wickramasinghe, P. Bhatnagar, J. Leal, R. Luengo-Fernandez, R. Burns, M. Rayner, and N. Townsend, "European cardiovascular disease statistics 2017," 2017.

[7] D. S. Amin and B.-R. Fethi, "Features for heartbeat sound signal normal and pathological," *Recent Patents on Computer Science*, vol. 1, no. 1, pp. 1–8, 2008.

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

[9] S. Reichert, R. Gass, C. Brandt, and E. Andrès, "Analysis of respiratory sounds: state of the art," *Clinical medicine. Circulatory, respiratory and pulmonary medicine*, vol. 2, pp. CCRPM–S530, 2008.

[10] S. S. Ching and Y. K. Tan, "Spectral analysis of bowel sounds in intestinal obstruction using an electronic stethoscope," *World Journal of Gastroenterology: WJG*, vol. 18, no. 33, p. 4585, 2012.

[11] K. Kumar and W. R. Thompson, "Evaluation of cardiac auscultation skills in pediatric residents," *Clinical pediatrics*, vol. 52, no. 1, pp. 66–73, 2013.

[12] M. Lam, T. Lee, P. Boey, W. Ng, H. Hey, K. Ho, and P. Cheong, "Factors influencing cardiac auscultation proficiency in physician trainees," *Singapore medical journal*, vol. 46, no. 1, p. 11, 2005.

[13] T. Flenady, T. Dwyer, and J. Applegarth, "Accurate respiratory rates count: So should you!," *Australasian Emergency Nursing Journal*, vol. 20, no. 1, pp. 45–47, 2017.

[14] S. Fleming, M. Thompson, R. Stevens, C. Heneghan, A. Plüddemann, I. Maconochie, L. Tarassenko, and D. Mant, "Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: a systematic review of observational studies," *The Lancet*, vol. 377, no. 9770, pp. 1011–1018, 2011.

[15] J. McFadden, R. Price, H. Eastwood, and R. Briggs, "Raised respiratory rate in elderly patients: a valuable physical sign.," *Br Med J (Clin Res Ed)*, vol. 284, no. 6316, pp. 626–627, 1982.

[16] S. B. Park and D. Khattar, "Tachypnea," 2019.

[17] W. H. Organization *et al.*, "Acute respiratory infections," tech. rep., World Health Organization, 1990.

[18] A. M. Chan, N. Ferdosi, and R. Narasimhan, "Ambulatory respiratory rate detection using ecg and a triaxial accelerometer," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 4058–4061, IEEE, 2013.

[19] S.-W. Deng and J.-Q. Han, "Towards heart sound classification without segmentation via autocorrelation feature and diffusion maps," *Future Generation Computer Systems*, vol. 60, pp. 13–21, 2016.

[20] H. Liang, S. Lukkarinen, and I. Hartimo, "Heart sound segmentation algorithm based on heart sound envelogram," in *Computers in Cardiology 1997*, pp. 105–108, IEEE, 1997.

[21] S. Kang, R. Doroshow, J. McConnaughey, A. Khandoker, and R. Shekhar, "Heart sound segmentation toward automated heart murmur classification in pediatric patents," in *2015 8th International Conference on Signal Processing, Image Processing and Pattern Recognition (SIP)*, pp. 9–12, IEEE, 2015.

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

[23] P. Wang, Y. Kim, L. Ling, and C. Soh, "First heart sound detection for phonocardiogram segmentation," in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, pp. 5519–5522, IEEE, 2006.

[24] J. Pedrosa, A. Castro, and T. T. Vinhoza, "Automatic heart sound segmentation and murmur detection in pediatric phonocardiograms," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 2294–2297, IEEE, 2014.

[25] F. D. L. Hedayioglu, “Heart sound segmentation for digital stethoscope integration,” 2011.

[26] N. Kouras, D. Boutana, and M. Benidir, “Wavelet based segmentation and time-frequency caracterisation of some abnormal heart sound signals,” in *2012 24th International Conference on Microelectronics (ICM)*, pp. 1–4, IEEE, 2012.

[27] D. Kumar, P. d. Carvalho, M. Antunes, J. Henriques, M. Maldonado, R. Schmidt, and J. Habetha, “Wavelet transform and simplicity based heart murmur segmentation,” in *2006 Computers in Cardiology*, pp. 173–176, IEEE, 2006.

[28] Z. Dokur and T. Ölmez, “Heart sound classification using wavelet transform and incremental self-organizing map,” *Digital Signal Processing*, vol. 18, no. 6, pp. 951–959, 2008.

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

[30] M. El-Segaier, O. Lilja, S. Lukkarinen, L. Sörnmo, R. Sepponen, and E. Pesonen, “Computer-based detection and analysis of heart sound and murmur,” *Annals of Biomedical Engineering*, vol. 33, no. 7, pp. 937–942, 2005.

[31] H. Liang, S. Lukkarinen, and I. Hartimo, “Heart sound segmentation algorithm based on heart sound envelogram,” in *Computers in Cardiology 1997*, pp. 105–108, IEEE, 1997.

[32] X. Wang, Y. Li, C. Sun, and C. Liu, “Detection of the first and second heart sound using heart sound energy,” in *2009 2nd International Conference on Biomedical Engineering and Informatics*, pp. 1–4, IEEE, 2009.

[33] S. Kang, R. Doroshow, J. McConnaughey, A. Khandoker, and R. Shekhar, “Heart sound segmentation toward automated heart murmur classification in pediatric patents,” in *2015 8th International Conference on Signal Processing, Image Processing and Pattern Recognition (SIP)*, pp. 9–12, IEEE, 2015.

[34] A. Iwata, N. Ishii, N. Suzumura, and K. Ikegaya, “Algorithm for detecting the first and the second heart sounds by spectral tracking,” *Medical and Biological Engineering and Computing*, vol. 18, no. 1, pp. 19–26, 1980.

[35] H. Liang, S. Lukkarinen, and I. Hartimo, “A boundary modification method for heart sound segmentation algorithm,” in *Computers in Cardiology 1998. Vol. 25 (Cat. No. 98CH36292)*, pp. 593–595, IEEE, 1998.

[36] T. Oskiper and R. Watrous, “Detection of the first heart sound using a time-delay neural network,” in *Computers in Cardiology*, pp. 537–540, IEEE, 2002.

[37] T. Ölmez and Z. Dokur, “Classification of heart sounds using an artificial neural network,” *Pattern Recognition Letters*, vol. 24, no. 1-3, pp. 617–629, 2003.

[38] D. Kumar, P. d. Carvalho, M. Antunes, J. Henriques, M. Maldonado, R. Schmidt, and J. Habetha, “Wavelet transform and simplicity based heart murmur segmentation,” in *2006 Computers in Cardiology*, pp. 173–176, IEEE, 2006.

[39] L. Zhong, X. Guo, A. Ji, and X. Ding, “A robust envelope extraction algorithm for cardiac sound signal segmentation,” in *2011 5th International Conference on Bioinformatics and Biomedical Engineering*, pp. 1–5, IEEE, 2011.

[40] A. D. Ricke, R. J. Povinelli, and M. T. Johnson, “Automatic segmentation of heart sound signals using hidden markov models,” in *Computers in Cardiology, 2005*, pp. 953–956, IEEE, 2005.

[41] C. S. Lima and M. J. Cardoso, “Phonocardiogram segmentation by using hidden markov models,” 2007.

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

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

[44] S. R. Bhatikar, C. DeGroff, and R. L. Mahajan, “A classifier based on the artificial neural network approach for cardiologic auscultation in pediatrics,” *Artificial intelligence in medicine*, vol. 33, no. 3, pp. 251–260, 2005.

[45] A. A. Sepehri, J. Hancq, T. Dutoit, A. Gharehbaghi, A. Kocharian, and A. Kiani, “Computerized screening of children congenital heart diseases,” *Computer methods and programs in biomedicine*, vol. 92, no. 2, pp. 186–192, 2008.

[46] A. Bouril, D. Aleinikava, M. S. Guillem, and G. M. Mirsky, “Automated classification of normal and abnormal heart sounds using support vector machines,” in *2016 Computing in Cardiology Conference (CinC)*, pp. 549–552, IEEE, 2016.

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

[48] L. Stankovic, M. Daković, and T. Thayaparan, *Time-frequency signal analysis with applications*. Artech house, 2014.

[49] A. M. Amiri and G. Armano, “An intelligent diagnostic system for congenital heart defects,” *Editorial Preface*, vol. 4, no. 7, pp. 93–98, 2013.

[50] X. Huai, S. Panote, D. Choi, and N. Kuwahara, “Heart sound recognition technology based on deep learning,” in *International Conference on Human-Computer Interaction*, pp. 491–500, Springer, 2020.

[51] J. Rubin, R. Abreu, A. Ganguli, S. Nelaturi, I. Matei, and K. Sricharan, “Recognizing abnormal heart sounds using deep learning,” *arXiv preprint arXiv:1707.04642*, 2017.

[52] F. Safara, S. Doraisamy, A. Azman, A. Jantan, and A. R. A. Ramaiah, “Multi-level basis selection of wavelet packet decomposition tree for heart sound classification,” *Computers in biology and medicine*, vol. 43, no. 10, pp. 1407–1414, 2013.

[53] E. J. Harfash, "Diagnostic the heart valve diseases using eigen vectors," *IJCSMC*, vol. 5, pp. 273–278, 2016.

[54] S. Ari and G. Saha, "In search of an optimization technique for artificial neural network to classify abnormal heart sounds," *Applied Soft Computing*, vol. 9, no. 1, pp. 330–340, 2009.

[55] B. Farzam and J. Shirazi, "The diagnosis of heart diseases based on pcg signals using mfcc coefficients and svm classifier," *IJISET-International Journal of Innovative Science, Engineering & Technology*, vol. 1, no. 10, 2014.

[56] R. Das, I. Turkoglu, and A. Sengur, "Diagnosis of valvular heart disease through neural networks ensembles," *Computer methods and programs in biomedicine*, vol. 93, no. 2, pp. 185–191, 2009.

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

[58] N. E. Singh-Miller and N. Singh-Miller, "Using spectral acoustic features to identify abnormal heart sounds," in *2016 Computing in Cardiology Conference (CinC)*, pp. 557–560, IEEE, 2016.

[59] J. Jaramillo-Garzon, A. Quiceno-Manrique, I. Godino-Llorente, and C. G. Castellanos-Dominguez, "Feature extraction for murmur detection based on support vector regression of time-frequency representations," in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1623–1626, IEEE, 2008.

[60] J. Liu, H. Wang, W. Liu, and J. Zhang, "Autonomous detection and classification of congenital heart disease using an auscultation vest," *Journal of Computational Information Systems*, vol. 8, no. 2, pp. 485–492, 2012.

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

[62] W.-C. Kao and C.-C. Wei, "Automatic phonocardiograph signal analysis for detecting heart valve disorders," *Expert Systems with Applications*, vol. 38, no. 6, pp. 6458–6468, 2011.

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

[64] W. Phanphaisarn, A. Roeksabutr, P. Wardkein, J. Koseeyaporn, and P. Yupapin, "Heart detection and diagnosis based on ecg and epcg relationships," *Medical devices (Auckland, NZ)*, vol. 4, p. 133, 2011.

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

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

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

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

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

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

[71] S. Banerjee and M. Mitra, “Application of cross wavelet transform for ecg pattern analysis and classification,” *IEEE transactions on instrumentation and measurement*, vol. 63, no. 2, pp. 326–333, 2013.

[72] S. Dilmac and M. Korurek, “Ecg heart beat classification method based on modified abc algorithm,” *Applied Soft Computing*, vol. 36, pp. 641–655, 2015.

[73] J. Mateo, A. Torres, A. Aparicio, and J. Santos, “An efficient method for ecg beat classification and correction of ectopic beats,” *Computers & Electrical Engineering*, vol. 53, pp. 219–229, 2016.

[74] S. Wahabi, S. Pouryayevali, S. Hari, and D. Hatzinakos, “On evaluating ecg biometric systems: Session-dependence and body posture,” *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 11, pp. 2002–2013, 2014.

[75] F. Agrafioti and D. Hatzinakos, “Ecg biometric analysis in cardiac irregularity conditions,” *Signal, Image and Video Processing*, vol. 3, no. 4, pp. 329–343, 2009.

[76] S.-C. Fang and H.-L. Chan, “Qrs detection-free electrocardiogram biometrics in the reconstructed phase space,” *Pattern Recognition Letters*, vol. 34, no. 5, pp. 595–602, 2013.

[77] R. G. Afkhami, G. Azarnia, and M. A. Tinati, “Cardiac arrhythmia classification using statistical and mixture modeling features of ecg signals,” *Pattern Recognition Letters*, vol. 70, pp. 45–51, 2016.

[78] Q. Li, C. Rajagopalan, and G. D. Clifford, “A machine learning approach to multi-level ecg signal quality classification,” *Computer methods and programs in biomedicine*, vol. 117, no. 3, pp. 435–447, 2014.

[79] R. J. Martis, U. R. Acharya, H. Prasad, C. K. Chua, and C. M. Lim, “Automated detection of atrial fibrillation using bayesian paradigm,” *Knowledge-Based Systems*, vol. 54, pp. 269–275, 2013.

[80] T. Ince, S. Kiranyaz, and M. Gabbouj, “A generic and robust system for automated patient-specific classification of ecg signals,” *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 5, pp. 1415–1426, 2009.

[81] Z. Zidelmal, A. Amirou, M. Adnane, and A. Belouchrani, “Qrs detection based on wavelet coefficients,” *Computer methods and programs in biomedicine*, vol. 107, no. 3, pp. 490–496, 2012.

[82] H.-Y. Lin, S.-Y. Liang, Y.-L. Ho, Y.-H. Lin, and H.-P. Ma, “Discrete-wavelet-transform-based noise removal and feature extraction for ecg signals,” *Irbm*, vol. 35, no. 6, pp. 351–361, 2014.

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

[84] J. Martinez, S. Olmos, and P. Laguna, “Evaluation of a wavelet-based ecg waveform detector on the qt database,” in *Computers in Cardiology 2000. Vol. 27 (Cat. 00CH37163)*, pp. 81–84, IEEE, 2000.

[85] M. Yochum, C. Renaud, and S. Jacquir, “Automatic detection of p, qrs and t patterns in 12 leads ecg signal based on cwt,” *Biomedical signal processing and control*, vol. 25, pp. 46–52, 2016.

[86] P. Karthikeyan, M. Murugappan, and S. Yaacob, “Ecg signal denoising using wavelet thresholding techniques in human stress assessment,” *International Journal on Electrical Engineering and Informatics*, vol. 4, no. 2, p. 306, 2012.

[87] S. Yazdani and J.-M. Vesin, “Extraction of qrs fiducial points from the ecg using adaptive mathematical morphology,” *Digital Signal Processing*, vol. 56, pp. 100–109, 2016.

[88] Y. Sun, K. L. Chan, and S. M. Krishnan, “Characteristic wave detection in ecg signal using morphological transform,” *BMC cardiovascular disorders*, vol. 5, no. 1, pp. 1–7, 2005.

[89] J. Pan and W. J. Tompkins, “A real-time qrs detection algorithm,” *IEEE Trans. Biomed. Eng*, vol. 32, no. 3, pp. 230–236, 1985.

[90] P. Phukpattaranont, “Qrs detection algorithm based on the quadratic filter,” *Expert Systems with Applications*, vol. 42, no. 11, pp. 4867–4877, 2015.

[91] J. A. Van Alste and T. Schilder, “Removal of base-line wander and power-line interference from the ecg by an efficient fir filter with a reduced number of taps,” *IEEE transactions on biomedical engineering*, no. 12, pp. 1052–1060, 1985.

[92] M. S. Manikandan and K. Soman, “A novel method for detecting r-peaks in electrocardiogram (ecg) signal,” *Biomedical Signal Processing and Control*, vol. 7, no. 2, pp. 118–128, 2012.

[93] J. N. Mak, Y. Hu, and K. D. Luk, “An automated ecg-artifact removal method for trunk muscle surface emg recordings,” *Medical engineering & physics*, vol. 32, no. 8, pp. 840–848, 2010.

[94] H. Zhu and J. Dong, "An r-peak detection method based on peaks of shannon energy envelope," *Biomedical Signal Processing and Control*, vol. 8, no. 5, pp. 466–474, 2013.

[95] H. Beyramienanlou and N. Lotfivand, "Shannon's energy based algorithm in ecg signal processing," *Computational and mathematical methods in medicine*, vol. 2017, 2017.

[96] A. Karimipour and M. R. Homaeinezhad, "Real-time electrocardiogram p-qrs-t detection–delineation algorithm based on quality-supported analysis of characteristic templates," *Computers in biology and medicine*, vol. 52, pp. 153–165, 2014.

[97] F. Bouaziz, D. Boutana, and M. Benidir, "Multiresolution wavelet-based qrs complex detection algorithm suited to several abnormal morphologies," *IET Signal Processing*, vol. 8, no. 7, pp. 774–782, 2014.

[98] A. H. Khandoker, M. Palaniswami, and C. K. Karmakar, "Support vector machines for automated recognition of obstructive sleep apnea syndrome from ecg recordings," *IEEE transactions on information technology in biomedicine*, vol. 13, no. 1, pp. 37–48, 2008.

[99] Y.-C. Yeh, W.-J. Wang, and C. W. Chiou, "Cardiac arrhythmia diagnosis method using linear discriminant analysis on ecg signals," *Measurement*, vol. 42, no. 5, pp. 778–789, 2009.

[100] Y.-C. Yeh, W.-J. Wang, and C. W. Chiou, "Feature selection algorithm for ecg signals using range-overlaps method," *Expert Systems with Applications*, vol. 37, no. 4, pp. 3499–3512, 2010.

[101] M. Korürek and B. Doğan, "Ecg beat classification using particle swarm optimization and radial basis function neural network," *Expert systems with Applications*, vol. 37, no. 12, pp. 7563–7569, 2010.

[102] P. Langley, E. J. Bowers, and A. Murray, "Principal component analysis as a tool for analyzing beat-to-beat changes in ecg features: application to ecg-derived respiration," *IEEE transactions on biomedical engineering*, vol. 57, no. 4, pp. 821–829, 2009.

[103] S. I. Safie, J. J. Soraghan, and L. Petropoulakis, "Electrocardiogram (ecg) biometric authentication using pulse active ratio (par)," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 4, pp. 1315–1322, 2011.

[104] M. R. Homaeinezhad, S. A. Atyabi, E. Tavakkoli, H. N. Toosi, A. Ghaffari, and R. Ebrahimpour, "Ecg arrhythmia recognition via a neuro-svm–knn hybrid classifier with virtual qrs image-based geometrical features," *Expert Systems with Applications*, vol. 39, no. 2, pp. 2047–2058, 2012.

[105] M. M. Tantawi, K. Revett, A. Salem, and M. F. Tolba, "Fiducial feature reduction analysis for electrocardiogram (ecg) based biometric recognition," *Journal of Intelligent Information Systems*, vol. 40, no. 1, pp. 17–39, 2013.

[106] E. Mazomenos, T. Chen, A. Acharyya, A. Bhattacharya, J. Rosengarten, and K. Maharatna, “A time-domain morphology and gradient based algorithm for ecg feature extraction,” in *2012 IEEE International conference on industrial technology*, pp. 117–122, IEEE, 2012.

[107] Q. Li, C. Rajagopalan, and G. D. Clifford, “Ventricular fibrillation and tachycardia classification using a machine learning approach,” *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 6, pp. 1607–1613, 2013.

[108] S.-M. Dima, C. Panagiotou, E. B. Mazomenos, J. A. Rosengarten, K. Maharatna, J. V. Gialelis, N. Curzen, and J. Morgan, “On the detection of myocadial scar based on ecg/vcg analysis,” *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 12, pp. 3399–3409, 2013.

[109] C. Song, K. Liu, X. Zhang, L. Chen, and X. Xian, “An obstructive sleep apnea detection approach using a discriminative hidden markov model from ecg signals,” *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 7, pp. 1532–1542, 2015.

[110] M. Javadi, S. A. A. A. Arani, A. Sajedin, and R. Ebrahimpour, “Classification of ecg arrhythmia by a modular neural network based on mixture of experts and negatively correlated learning,” *Biomedical Signal Processing and Control*, vol. 8, no. 3, pp. 289–296, 2013.

[111] F. A. Elhaj, N. Salim, A. R. Harris, T. T. Swee, and T. Ahmed, “Arrhythmia recognition and classification using combined linear and nonlinear features of ecg signals,” *Computer methods and programs in biomedicine*, vol. 127, pp. 52–63, 2016.

[112] S. Ergin, A. K. Uysal, E. S. Gunal, S. Gunal, and M. B. Gulmezoglu, “Ecg based biometric authentication using ensemble of features,” in *2014 9th Iberian Conference on Information Systems and Technologies (CISTI)*, pp. 1–6, IEEE, 2014.

[113] S. Gunal, S. Ergin, E. S. Gunal, and A. K. Uysal, “Ecg classification using ensemble of features,” in *2013 47th Annual Conference on Information Sciences and Systems (CISS)*, pp. 1–5, IEEE, 2013.

[114] Y. Özbay, “A new approach to detection of ecg arrhythmias: Complex discrete wavelet transform based complex valued artificial neural network,” *Journal of Medical Systems*, vol. 33, no. 6, pp. 435–445, 2009.

[115] H. Khorrami and M. Moavenian, “A comparative study of dwt, cwt and dct transformations in ecg arrhythmias classification,” *Expert systems with Applications*, vol. 37, no. 8, pp. 5751–5757, 2010.

[116] E. D. Übeyli, “Statistics over features: Eeg signals analysis,” *Computers in Biology and Medicine*, vol. 39, no. 8, pp. 733–741, 2009.

[117] W. Karlen, C. Mattiussi, and D. Floreano, “Sleep and wake classification with ecg and respiratory effort signals,” *IEEE Transactions on Biomedical Circuits and Systems*, vol. 3, no. 2, pp. 71–78, 2009.

[118] M. S. Islam and N. Alajlan, “A morphology alignment method for resampled heartbeat signals,” *Biomedical Signal Processing and Control*, vol. 8, no. 3, pp. 315–324, 2013.

[119] J. Huang, B. Chen, B. Yao, and W. He, “Ecg arrhythmia classification using stft-based spectrogram and convolutional neural network,” *IEEE access*, vol. 7, pp. 92871–92880, 2019.

[120] A. Diker, Z. Cömert, E. Avcı, M. Toğaçar, and B. Ergen, “A novel application based on spectrogram and convolutional neural network for ecg classification,” in *2019 1st International Informatics and Software Engineering Conference (UBMYK)*, pp. 1–6, IEEE, 2019.

[121] J. McNames and A. Fraser, “Obstructive sleep apnea classification based on spectrogram patterns in the electrocardiogram,” in *Computers in Cardiology 2000. vol. 27 (Cat. 00CH37163)*, pp. 749–752, IEEE, 2000.

[122] V. Gupta, M. Mittal, V. Mittal, and A. Gupta, “Ecg signal analysis using cwt, spectrogram and autoregressive technique,” *Iran Journal of Computer Science*, vol. 4, no. 4, pp. 265–280, 2021.

[123] Y. Kutlu and D. Kuntalp, “A multi-stage automatic arrhythmia recognition and classification system,” *Computers in biology and medicine*, vol. 41, no. 1, pp. 37–45, 2011.

[124] Y. Kutlu and D. Kuntalp, “Feature extraction for ecg heartbeats using higher order statistics of wpd coefficients,” *Computer methods and programs in biomedicine*, vol. 105, no. 3, pp. 257–267, 2012.

[125] M. Seera, C. P. Lim, W. S. Liew, E. Lim, and C. K. Loo, “Classification of electrocardiogram and auscultatory blood pressure signals using machine learning models,” *Expert Systems with Applications*, vol. 42, no. 7, pp. 3643–3652, 2015.

[126] C. Bruser, J. Diesel, M. D. Zink, S. Winter, P. Schauerte, and S. Leonhardt, “Automatic detection of atrial fibrillation in cardiac vibration signals,” *IEEE journal of biomedical and health informatics*, vol. 17, no. 1, pp. 162–171, 2012.

[127] Q. A. Rahman, L. G. Tereshchenko, M. Kongkatong, T. Abraham, M. R. Abraham, and H. Shatkay, “Utilizing ecg-based heartbeat classification for hypertrophic cardiomyopathy identification,” *IEEE transactions on nanobioscience*, vol. 14, no. 5, pp. 505–512, 2015.

[128] T. Chen, E. B. Mazomenos, K. Maharatna, S. Dasmahapatra, and M. Niranjan, “Design of a low-power on-body ecg classifier for remote cardiovascular monitoring systems,” *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 3, no. 1, pp. 75–85, 2013.

[129] A. Jovic and N. Bogunovic, “Electrocardiogram analysis using a combination of statistical, geometric, and nonlinear heart rate variability features,” *Artificial intelligence in medicine*, vol. 51, no. 3, pp. 175–186, 2011.

[130] J. H. Abawajy, A. V. Kelarev, and M. Chowdhury, “Multistage approach for clustering and classification of ecg data,” *Computer methods and programs in biomedicine*, vol. 112, no. 3, pp. 720–730, 2013.

[131] S.-N. Yu and K.-T. Chou, “Selection of significant independent components for ecg beat classification,” *Expert Systems with Applications*, vol. 36, no. 2, pp. 2088–2096, 2009.

[132] U. Orhan, "Real-time chf detection from ecg signals using a novel discretization method," *Computers in biology and medicine*, vol. 43, no. 10, pp. 1556–1562, 2013.

[133] H. Yang, C. Kan, G. Liu, and Y. Chen, "Spatiotemporal differentiation of myocardial infarctions," *IEEE Transactions on Automation Science and Engineering*, vol. 10, no. 4, pp. 938–947, 2013.

[134] N. F. Güler, E. D. Übeyli, and I. Güler, "Recurrent neural networks employing lyapunov exponents for eeg signals classification," *Expert systems with applications*, vol. 29, no. 3, pp. 506–514, 2005.

[135] S. Dutta, A. Chatterjee, and S. Munshi, "Identification of ecg beats from cross-spectrum information aided learning vector quantization," *Measurement*, vol. 44, no. 10, pp. 2020–2027, 2011.

[136] S. Singh, S. K. Pandey, U. Pawar, and R. R. Janghel, "Classification of ecg arrhythmia using recurrent neural networks," *Procedia computer science*, vol. 132, pp. 1290–1297, 2018.

[137] R. Ceylan, Y. Özbay, and B. Karlik, "A novel approach for classification of ecg arrhythmias: Type-2 fuzzy clustering neural network," *Expert Systems with Applications*, vol. 36, no. 3, pp. 6721–6726, 2009.

[138] H. H. Haseena, A. T. Mathew, and J. K. Paul, "Fuzzy clustered probabilistic and multi layered feed forward neural networks for electrocardiogram arrhythmia classification," *Journal of Medical Systems*, vol. 35, no. 2, pp. 179–188, 2011.

[139] Y. Özbay, R. Ceylan, and B. Karlik, "Integration of type-2 fuzzy clustering and wavelet transform in a neural network based ecg classifier," *Expert Systems with Applications*, vol. 38, no. 1, pp. 1004–1010, 2011.

[140] I. Nejadgholi, M. H. Moradi, and F. Abdolali, "Using phase space reconstruction for patient independent heartbeat classification in comparison with some benchmark methods," *Computers in biology and medicine*, vol. 41, no. 6, pp. 411–419, 2011.

[141] P. Li, Y. Wang, J. He, L. Wang, Y. Tian, T.-s. Zhou, T. Li, and J.-s. Li, "High-performance personalized heartbeat classification model for long-term ecg signal," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 1, pp. 78–86, 2016.

[142] L. Brochard, G. S. Martin, L. Blanch, P. Pelosi, F. J. Belda, A. Jubran, L. Gattinoni, J. Mancebo, V. M. Ranieri, J.-C. M. Richard, *et al.*, "Clinical review: Respiratory monitoring in the icu-a consensus of 16," *Critical Care*, vol. 16, no. 2, pp. 1–14, 2012.

[143] I. Wheatley, "Respiratory rate 3: How to take an accurate measurement," *Nursing Times*, vol. 114, no. 7, pp. 21–22, 2018.

[144] P. H. Charlton, M. Villarroel, and F. Salguiero, "Waveform analysis to estimate respiratory rate," *Secondary Analysis of Electronic Health Records*, pp. 377–390, 2016.

[145] J. M. Ansermino, G. Dumont, and A. S. Ginsburg, "How uncertain is our reference standard for respiratory rate measurement?," *American journal of respiratory and critical care medicine*, vol. 199, no. 8, pp. 1036–1037, 2019.

[146] S. Yabuki, H. Toyama, Y. Takei, T. Wagatsuma, H. Yabuki, and M. Yamauchi, "Influences of environmental noise level and respiration rate on the accuracy of acoustic respiration rate monitoring," *Journal of Clinical Monitoring and Computing*, vol. 32, no. 1, pp. 127–132, 2018.

[147] W. Einthoven, G. Fahr, and A. De Waart, "On the direction and manifest size of the variations of potential in the human heart and on the influence of the position of the heart on the form of the electrocardiogram," *American heart journal*, vol. 40, no. 2, pp. 163–211, 1950.

[148] J. T. Flaherty, S. D. Blumenschein, A. W. Alexander, R. D. Gentzler, T. M. Gallie, J. P. Boineau, and M. S. Spach, "Influence of respiration on recording cardiac potentials: Isopotential surface-mapping and vectorcardiographic studies," *The American journal of cardiology*, vol. 20, no. 1, pp. 21–28, 1967.

[149] H. Riekkinen and P. Rautaharju, "Body position, electrode level, and respiration effects on the frank lead electrocardiogram.," *Circulation*, vol. 53, no. 1, pp. 40–45, 1976.

[150] A. Travaglini, C. Lamberti, J. DeBie, and M. Ferri, "Respiratory signal derived from eight-lead ecg," in *Computers in Cardiology 1998. Vol. 25 (Cat. No. 98CH36292)*, pp. 65–68, IEEE, 1998.

[151] J. Boyle, N. Bidargaddi, A. Sarela, and M. Karunanithi, "Automatic detection of respiration rate from ambulatory single-lead ecg," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 6, pp. 890–896, 2009.

[152] F. Pinciroli, R. Rossi, and L. Vergani, "Detection of electrical axis variation for the extraction of respiratory information," *Computers in Cardiology*, vol. 2, pp. 499–502, 1985.

[153] B. Mazzanti, C. Lamberti, and J. De Bie, "Validation of an ecg-derived respiration monitoring method," in *Computers in Cardiology, 2003*, pp. 613–616, IEEE, 2003.

[154] L. Zhao, S. Reisman, and T. Findley, *Derivation of respiration from electrocardiogram during heart rate variability studies*. IEEE, 1994.

[155] K. Behbehani, S. Vijendra, J. Burk, and E. Lucas, "An investigation of the mean electrical axis angle and respiration during sleep," in *Proceedings of the Second Joint 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society][Engineering in Medicine and Biology*, vol. 2, pp. 1550–1551, IEEE, 2002.

[156] D. Dobrev and I. Daskalov, "Two-electrode telemetric instrument for infant heart rate and apnea monitoring," *Medical engineering & physics*, vol. 20, no. 10, pp. 729–734, 1999.

[157] M. Sakai, R. Sekine, and X. Zhu, "Single-channel ecg suitable for ecg-derived respiration," *Biomedical Physics & Engineering Express*, vol. 5, no. 5, p. 055002, 2019.

[158] L. Mason, *Signal processing methods for non-invasive respiration monitoring*. University of Oxford Oxford, 2002.

[159] D. Widjaja, J. Taelman, S. Vandeput, M. A. Braeken, R. A. Otte, B. R. Van den Bergh, and S. Van Huffel, "Ecg-derived respiration: comparison and new measures for respiratory variability," in *2010 Computing in Cardiology*, pp. 149–152, IEEE, 2010.

[160] L. S. Correa, E. Laciar, A. Torres, and R. Jane, "Performance evaluation of three methods for respiratory signal estimation from the electrocardiogram," in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 4760–4763, IEEE, 2008.

[161] P. De Chazal, C. Heneghan, E. Sheridan, R. Reilly, P. Nolan, and M. O'Malley, "Automated processing of the single-lead electrocardiogram for the detection of obstructive sleep apnoea," *IEEE transactions on biomedical engineering*, vol. 50, no. 6, pp. 686–696, 2003.

[162] M. Campolo, D. Labate, F. La Foresta, F. Morabito, A. Lay-Ekuakille, and P. Vergallo, "Ecg-derived respiratory signal using empirical mode decomposition," in *2011 IEEE International Symposium on Medical Measurements and Applications*, pp. 399–403, IEEE, 2011.

[163] R. Auer, D. C. Bauer, P. Marques-Vidal, J. Butler, L. J. Min, J. Cornuz, S. Satterfield, A. B. Newman, E. Vittinghoff, N. Rodondi, *et al.*, "Association of major and minor ecg abnormalities with coronary heart disease events," *Jama*, vol. 307, no. 14, pp. 1497–1505, 2012.

[164] J. Wartak, *Phonocardiology; integrated study of heart sounds and murmurs*. HarperCollins Publishers, 1972.

[165] H. Shino, H. Yoshida, K. Yana, K. Harada, J. Sudoh, and E. Harasewa, "Detection and classification of systolic murmur for phonocardiogram screening," in *Proceedings of 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 1, pp. 123–124, IEEE, 1996.

[166] Z. Syed, J. Guttag, R. Levine, F. Nesta, and D. Curtis, "Automated auscultation system," Dec. 23 2004. US Patent App. 10/464,267.

[167] C. Ahlström, *Processing of the Phonocardiographic Signal: methods for the intelligent stethoscope*. PhD thesis, Institutionen för medicinsk teknik, 2006.

[168] A. Andresen, P. Galen, R. Warner, and R. Selvester, "Combined ecg and sound chart report and methodology," June 22 2006. US Patent App. 11/140,010.

[169] M. Homaeinezhad, P. Sabetian, A. Feizollahi, A. Ghaffari, and R. Rahmani, "Parametric modelling of cardiac system multiple measurement signals: an open-source computer framework for performance evaluation of ecg, pcg and abp event detectors," *Journal of medical engineering & technology*, vol. 36, no. 2, pp. 117–134, 2012.

[170] M. Zarrabi, H. Parsaei, R. Boostani, A. Zare, Z. Dorfeshan, K. Zarrabi, and J. Kojuri, "A system for accurately predicting the risk of myocardial infarction using pcg, ecg and clinical features," *Biomedical Engineering: Applications, Basis and Communications*, vol. 29, no. 03, p. 1750023, 2017.

[171] M. Okada, “Chest wall maps of heart sounds and murmurs,” *Computers and biomedical research*, vol. 15, no. 3, pp. 281–294, 1982.

[172] M. Cozic, L.-G. Durand, and R. Guardo, “Development of a cardiac acoustic mapping system,” *Medical and Biological Engineering and Computing*, vol. 36, no. 4, pp. 431–437, 1998.

[173] Y. Bahadirlar and H. Ö. Gülçür, “Cardiac passive acoustic localization: Cardiopal,” *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 6, no. 3, pp. 243–260, 2000.

[174] F. Nogata, Y. Yokota, Y. Kawamura, H. Morita, and Y. Uno, “Novel technique for visualizing heart motion without using ultrasonic cardiography,” in *Proceedings of 2012 IEEE-EMBS International Conference on Biomedical and Health Informatics*, pp. 249–252, IEEE, 2012.

[175] C. Sapsanis, N. Welsh, M. Pozin, G. Garreau, G. Tognetti, H. Bakhshaee, P. O. Pouliquen, R. Mitral, W. R. Thompson, and A. G. Andreou, “Stethovest: A simultaneous multichannel wearable system for cardiac acoustic mapping,” in *2018 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pp. 1–4, IEEE, 2018.

[176] M. Kania, H. Rix, M. Fereniec, H. Zavala-Fernandez, D. Janusek, T. Mroczka, G. Stix, and R. Maniewski, “The effect of precordial lead displacement on ecg morphology,” *Medical & biological engineering & computing*, vol. 52, no. 2, pp. 109–119, 2014.

[177] D. Ricciardi, I. Cavallari, A. Creta, G. D. Giovanni, V. Calabrese, N. D. Belardino, S. Mega, I. Colaiori, L. Ragni, C. Proscia, A. Nenna, and G. D. Sciascio, “Impact of the high-frequency cutoff of bandpass filtering on ecg quality and clinical interpretation: A comparison between 40hz and 150hz cutoff in a surgical preoperative adult outpatient population,” *Journal of Electrocardiology*, vol. 49, no. 5, pp. 691 – 695, 2016.

[178] P. Laguna, R. Jané, and P. Caminal, “Adaptive filtering of ecg baseline wander,” in *1992 14th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2, pp. 508–509, Oct 1992.

[179] J. H. O’Keefe Jr, S. C. Hammill, M. S. Freed, and S. M. Pogwizd, *The complete guide to ECGs*. Jones & Bartlett Publishers, 2010.

[180] B. Karnath and W. Thornton, “Auscultation of the heart,” *Hospital Physician*, vol. 38, no. 9, pp. 39–45, 2002.

[181] T. Daniela, “Clinical characteristics and prognosis significance of bundle-branch block (bbb) associated with acute myocardial infarction (ami).,” *Rom J Intern Med*, vol. 34, no. 3-4, pp. 211–215, 1996.

[182] D. S. Gerbarg, A. Taranta, M. Spagnuolo, and J. J. Hofler, “Computer analysis of phonocardiograms,” *Progress in Cardiovascular Diseases*, vol. 5, no. 4, pp. 393–405, 1963.

[183] W. Phanphaisarn, A. Roeksabutr, P. Wardkein, J. Koseeyaporn, and P. Yupapin, “Heart detection and diagnosis based on ecg and epcg relationships,” *Medical devices (Auckland, NZ)*, vol. 4, p. 133, 2011.

[184] A. Zeyer, P. Doetsch, P. Voigtlaender, R. Schlüter, and H. Ney, "A comprehensive study of deep bidirectional LSTM RNNs for acoustic modeling in speech recognition," in *2017 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pp. 2462–2466, IEEE, 2017.

[185] G. Liu and J. Guo, "Bidirectional LSTM with attention mechanism and convolutional layer for text classification," *Neurocomputing*, vol. 337, pp. 325–338, 2019.

[186] F. Demir, A. Şengür, V. Bajaj, and K. Polat, "Towards the classification of heart sounds based on convolutional deep neural network," *Health information science and systems*, vol. 7, no. 1, pp. 1–9, 2019.

[187] X. Cheng, J. Huang, Y. Li, and G. Gui, "Design and application of a laconic heart sound neural network," *IEEE Access*, vol. 7, pp. 124417–124425, 2019.

[188] B. Ergen, Y. Tatar, and H. O. Gulcur, "Time-frequency analysis of phonocardiogram signals using wavelet transform: a comparative study," *Computer methods in biomechanics and biomedical engineering*, vol. 15, no. 4, pp. 371–381, 2012.

[189] S. S. Mahmoud, Z. M. Hussain, I. Cosic, and Q. Fang, "Time-frequency analysis of normal and abnormal biological signals," *Biomedical Signal Processing and Control*, vol. 1, no. 1, pp. 33–43, 2006.

[190] M. Obaidat, "Phonocardiogram signal analysis: techniques and performance comparison," *Journal of medical engineering & technology*, vol. 17, no. 6, pp. 221–227, 1993.

[191] Z. Peng, G. Meng, F. Chu, Z. Lang, W. Zhang, and Y. Yang, "Polynomial chirplet transform with application to instantaneous frequency estimation," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 9, pp. 3222–3229, 2011.

[192] F. Auger, P. Flandrin, P. Gonçalvès, and O. Lemoine, "Time-frequency toolbox," *CNRS France-Rice University*, vol. 46, 1996.

[193] L. H. Cherif, S. Debbal, and F. Bereksi-Reguig, "Choice of the wavelet analyzing in the phonocardiogram signal analysis using the discrete and the packet wavelet transform," *Expert Systems with Applications*, vol. 37, no. 2, pp. 913–918, 2010.

[194] P. Vikhe, S. Hamde, and N. Nehe, "Wavelet transform based abnormality analysis of heart sound," in *2009 International Conference on Advances in Computing, Control, and Telecommunication Technologies*, pp. 367–371, IEEE, 2009.

[195] S. Debbal and F. Bereksi-Reguig, "Analysis of the second heart sound using continuous wavelet transform," *Journal of medical engineering & technology*, vol. 28, no. 4, pp. 151–156, 2004.

[196] A. Taebi and H. A. Mansy, "Analysis of seismocardiographic signals using polynomial chirplet transform and smoothed pseudo Wigner-Ville distribution," in *2017 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, pp. 1–6, IEEE, 2017.

[197] S. Mann and S. Haykin, "The chirplet transform: Physical considerations," *IEEE Transactions on Signal Processing*, vol. 43, no. 11, pp. 2745–2761, 1995.

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

[199] A. Djebbari and F. Bereksi-Reguig, "Detection of the valvular split within the second heart sound using the reassigned smoothed pseudo Wigner-Ville distribution," *Biomedical engineering online*, vol. 12, no. 1, pp. 1–21, 2013.

[200] T. Chen, L. Xiang, and M. Zhang, "Recognition of heart sound based on distribution of Choi-Williams," *Research on Biomedical Engineering*, vol. 31, pp. 189–195, 2015.

[201] A. Taebi and H. A. Mansy, "Time-frequency distribution of seismocardiographic signals: A comparative study," *Bioengineering*, vol. 4, no. 2, p. 32, 2017.

[202] J. Cui and D. Wang, "Biosignal analysis with matching-pursuit based adaptive chirplet transform," *arXiv preprint arXiv:1709.08328*, 2017.

[203] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, no. 3, pp. 211–252, 2015.

[204] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, *et al.*, "Searching for MobilenetV3," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1314–1324, 2019.

[205] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[206] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7132–7141, 2018.

[207] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017.

[208] M. Tan, B. Chen, R. Pang, V. Vasudevan, M. Sandler, A. Howard, and Q. V. Le, "MNasNet: Platform-aware neural architecture search for mobile," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2820–2828, 2019.

[209] T.-J. Yang, A. Howard, B. Chen, X. Zhang, A. Go, M. Sandler, V. Sze, and H. Adam, "Netadapt: Platform-aware neural network adaptation for mobile applications," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 285–300, 2018.

[210] M. Nabih-Ali, E.-S. A. El-Dahshan, and A. S. Yahia, "A review of intelligent systems for heart sound signal analysis," *Journal of medical engineering & technology*, vol. 41, no. 7, pp. 553–563, 2017.

[211] V. Arora, R. Leekha, R. Singh, and I. Chana, “Heart sound classification using machine learning and phonocardiogram,” *Modern Physics Letters B*, vol. 33, no. 26, p. 1950321, 2019.

[212] V. Sujadevi, K. Soman, R. Vinayakumar, and A. Prem Sankar, “Anomaly detection in phonocardiogram employing deep learning,” in *Computational Intelligence in Data Mining*, pp. 525–534, Springer, 2019.

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

[214] G. D. Clifford, C. Liu, B. E. Moody, J. M. Roig, S. E. Schmidt, Q. Li, I. Silva, and R. G. Mark, “Recent advances in heart sound analysis,” *Physiological measurement*, vol. 38, pp. E10–E25, 2017.

[215] T.-c. I. Yang and H. Hsieh, “Classification of acoustic physiological signals based on deep learning neural networks with augmented features,” in *2016 Computing in Cardiology Conference (CinC)*, pp. 569–572, IEEE, 2016.

[216] C. Ittichaichareon, S. Suksri, and T. Yingthawornsuk, “Speech recognition using mfcc,” in *International conference on computer graphics, simulation and modeling*, vol. 9, 2012.

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

[218] Z. Ren, N. Cummins, V. Pandit, J. Han, K. Qian, and B. Schuller, “Learning image-based representations for heart sound classification,” in *Proceedings of the 2018 International Conference on Digital Health*, pp. 143–147, 2018.

[219] T. Nilanon, J. Yao, J. Hao, S. Purushotham, and Y. Liu, “Normal/abnormal heart sound recordings classification using convolutional neural network,” in *2016 computing in cardiology conference (CinC)*, pp. 585–588, IEEE, 2016.

[220] A. Rizal, A. Adz-Dzikri, M. Arik, and G. Fauzi, “Classification of normal and abnormal heart sound using continuous wavelet transform and ResNet-50,” *Technology Reports of Kansai University*, vol. 62, pp. 2595–2601, 06 2020.

[221] H. Alaskar, N. Alzhrani, A. Hussain, and F. Almarshed, “The implementation of pretrained AlexNet on PCG classification,” in *International Conference on Intelligent Computing*, pp. 784–794, Springer, 2019.

[222] S.-Y. Jung, C.-H. Liao, Y.-S. Wu, S.-M. Yuan, and C.-T. Sun, “Efficiently classifying lung sounds through depthwise separable CNN models with fused STFT and MFCC features,” *Diagnostics*, vol. 11, no. 4, p. 732, 2021.

[223] L. Zhang, D. Wang, C. Bao, Y. Wang, and K. Xu, “Large-scale whale-call classification by transfer learning on multi-scale waveforms and time-frequency features,” *Applied Sciences*, vol. 9, no. 5, p. 1020, 2019.

[224] I. McLoughlin, Z. Xie, Y. Song, H. Phan, and R. Palaniappan, "Time-frequency feature fusion for noise robust audio event classification," *Circuits, Systems, and Signal Processing*, vol. 39, no. 3, pp. 1672–1687, 2020.

[225] M. Jalayer, C. Orsenigo, and C. Vercellis, "Fault detection and diagnosis for rotating machinery: A model based on convolutional LSTM, fast fourier and continuous wavelet transforms," *Computers in Industry*, vol. 125, p. 103378, 2021.

[226] X. Liu, L. Lv, Y. Shen, P. Xiong, J. Yang, and J. Liu, "Multiscale space-time-frequency feature-guided multitask learning CNN for motor imagery EEG classification," *Journal of Neural Engineering*, vol. 18, no. 2, p. 026003, 2021.

[227] G. Liu, X. Han, L. Tian, W. Zhou, and H. Liu, "ECG quality assessment based on hand-crafted statistics and deep-learned S-transform spectrogram features," *Computer Methods and Programs in Biomedicine*, vol. 208, p. 106269, 2021.

[228] X. Yan and M. Jia, "A novel optimized SVM classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing," *Neurocomputing*, vol. 313, pp. 47–64, 2018.

[229] X. Bao, Y. Xu, and E. N. Kamavuako, "The effect of signal duration on the classification of heart sounds: A deep learning approach," *Sensors*, vol. 22, no. 6, p. 2261, 2022.

[230] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, 2019.

[231] R. Wightman, "Pytorch image models." https://github.com/rwightman/pytorch-image-models, 2019.

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

[233] X. Bao, Y. Xu, H.-K. Lam, M. Trabelsi, I. Chihi, L. Sidhom, and E. N. Kamavuako, "Time-frequency distributions of heart sound signals: A comparative study using convolutional neural networks," *arXiv preprint arXiv:2208.03128*, 2022.

[234] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7794–7803, 2018.

[235] M. A. Reyna, Y. Kiarashi, A. Elola, J. Oliveira, F. Renna, A. Gu, E. A. Perez-Alday, N. Sadr, A. Sharma, S. Mattos, *et al.*, "Heart murmur detection from phonocardiogram recordings: The George B. Moody PhysioNet Challenge 2022," *medRxiv*, 2022.

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

[237] A. Tao, K. Sapra, and B. Catanzaro, “Hierarchical multi-scale attention for semantic segmentation,” *arXiv preprint arXiv:2005.10821*, 2020.

[238] W. Shen, M. Zhou, F. Yang, C. Yang, and J. Tian, “Multi-scale convolutional neural networks for lung nodule classification,” in *International conference on information processing in medical imaging*, pp. 588–599, Springer, 2015.

[239] J. Kornej, C. S. Börschel, E. J. Benjamin, and R. B. Schnabel, “Epidemiology of atrial fibrillation in the 21st century: novel methods and new insights,” *Circulation research*, vol. 127, no. 1, pp. 4–20, 2020.

[240] C. T. January, L. S. Wann, J. S. Alpert, H. Calkins, J. E. Cigarroa, J. C. Cleveland, J. B. Conti, P. T. Ellinor, M. D. Ezekowitz, M. E. Field, *et al.*, “2014 aha/acc/hrs guideline for the management of patients with atrial fibrillation: a report of the american college of cardiology/american heart association task force on practice guidelines and the heart rhythm society,” *Journal of the American College of Cardiology*, vol. 64, no. 21, pp. e1–e76, 2014.

[241] E. Lee, E.-K. Choi, K.-D. Han, H. Lee, W.-S. Choe, S.-R. Lee, M.-J. Cha, W.-H. Lim, Y.-J. Kim, and S. Oh, “Mortality and causes of death in patients with atrial fibrillation: a nationwide population-based study,” *PLoS One*, vol. 13, no. 12, p. e0209687, 2018.

[242] X. Zhou, H. Ding, W. Wu, and Y. Zhang, “A real-time atrial fibrillation detection algorithm based on the instantaneous state of heart rate,” *PloS one*, vol. 10, no. 9, p. e0136544, 2015.

[243] E. De Giovanni, A. Aminifar, A. Luca, S. Yazdani, J.-M. Vesin, and D. Atienza, “A patient-specific methodology for prediction of paroxysmal atrial fibrillation onset,” in *2017 Computing in Cardiology (CinC)*, pp. 1–4, IEEE, 2017.

[244] V. Kalidas and L. S. Tamil, “Detection of atrial fibrillation using discrete-state markov models and random forests,” *Computers in biology and medicine*, vol. 113, p. 103386, 2019.

[245] B. Pourbabaee, M. J. Roshtkhari, and K. Khorasani, “Deep convolutional neural networks and learning ecg features for screening paroxysmal atrial fibrillation patients,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 12, pp. 2095–2104, 2018.

[246] A. Annavarapu and P. Kora, “Ecg-based atrial fibrillation detection using different orderings of conjugate symmetric–complex hadamard transform,” *International Journal of the Cardiovascular Academy*, vol. 2, no. 3, pp. 151–154, 2016.

[247] A. Rizwan, A. Zoha, I. B. Mabrouk, H. M. Sabbour, A. S. Al-Sumaiti, A. Alomainy, M. A. Imran, and Q. H. Abbasi, “A review on the state of the art in atrial fibrillation detection enabled by machine learning,” *IEEE reviews in biomedical engineering*, vol. 14, pp. 219–239, 2020.

[248] C. Liu, J. Oster, E. Reinertsen, Q. Li, L. Zhao, S. Nemati, and G. D. Clifford, “A comparison of entropy approaches for af discrimination,” *Physiological measurement*, vol. 39, no. 7, p. 074002, 2018.

[249] Z. Xiong, M. K. Stiles, and J. Zhao, "Robust ecg signal classification for detection of atrial fibrillation using a novel neural network," in *2017 Computing in Cardiology (CinC)*, pp. 1–4, IEEE, 2017.

[250] G. Petmezas, K. Haris, L. Stefanopoulos, V. Kilintzis, A. Tzavelis, J. A. Rogers, A. K. Katsaggelos, and N. Maglaveras, "Automated atrial fibrillation detection using a hybrid cnn-lstm network on imbalanced ecg datasets," *Biomedical Signal Processing and Control*, vol. 63, p. 102194, 2021.

[251] Y. Ping, C. Chen, L. Wu, Y. Wang, and M. Shu, "Automatic detection of atrial fibrillation based on cnn-lstm and shortcut connection," in *Healthcare*, vol. 8, p. 139, Multidisciplinary Digital Publishing Institute, 2020.

[252] G. F. Michaud and W. G. Stevenson, "Atrial fibrillation," *N Engl J Med.*, vol. 384(4), pp. 353–361, 2021.

[253] X. Wang, C. Ma, X. Zhang, H. Gao, G. Clifford, and C. Liu, "Paroxysmal atrial fibrillation events detection from dynamic ecg recordings: The 4th china physiological signal challenge 2021," in *Proc. PhysioNet*, pp. 1–83, 2021.

[254] A. Graves, S. Fernández, and J. Schmidhuber, "Bidirectional lstm networks for improved phoneme classification and recognition," in *International conference on artificial neural networks*, pp. 799–804, Springer, 2005.

[255] R. Kitney and O. Rompelman, *The study of heart-rate variability.* No. 3, Oxford University Press, USA, 1980.

[256] X. Bao, M. Howard, I. K. Niazi, and E. N. Kamavuako, "Comparison between embroidered and gel electrodes on ecg-derived respiration rate," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 2622–2625, IEEE, 2020.

[257] A. Shafti, R. B. R. Manero, A. M. Borg, K. Althoefer, and M. J. Howard, "Embroidered electromyography: A systematic design guide," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 9, pp. 1472–1480, 2016.

[258] S. Pitou, F. Wu, A. Shafti, B. Michael, R. Stopforth, and M. Howard, "Embroidered electrodes for control of affordable myoelectric prostheses," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1812–1817, IEEE, 2018.

[259] H. Zhang, L. Tian, L. Zhang, and G. Li, "Using textile electrode emg for prosthetic movement identification in transradial amputees," in *2013 IEEE International Conference on Body Sensor Networks*, pp. 1–5, IEEE, 2013.

[260] J. Löfhede, F. Seoane, and M. Thordstein, "Textile electrodes for eeg recording—a pilot study," *Sensors*, vol. 12, no. 12, pp. 16907–16919, 2012.

[261] T. Pola and J. Vanhala, "Textile electrodes in ecg measurement," in *2007 3rd International Conference on Intelligent Sensors, Sensor Networks and Information*, pp. 635–639, IEEE, 2007.

[262] C.-L. Shen, T.-H. Huang, P.-C. Hsu, Y.-C. Ko, F.-L. Chen, W.-C. Wang, T. Kao, and C.-T. Chan, “Respiratory rate estimation by using ecg, impedance, and motion sensing in smart clothing,” *Journal of Medical and Biological Engineering*, vol. 37, no. 6, pp. 826–842, 2017.

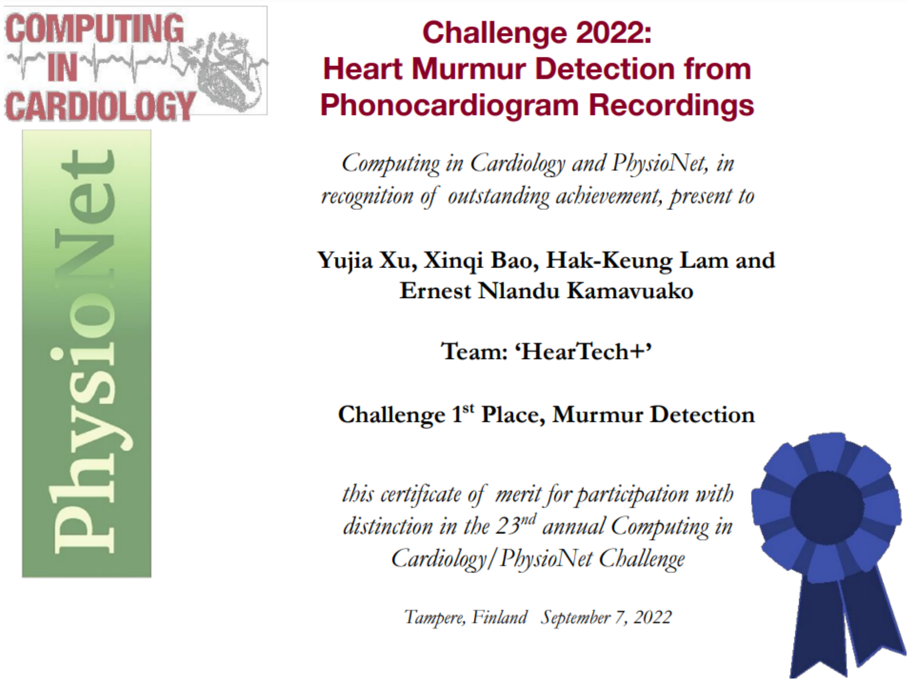
[263] V. Marozas, A. Petrenas, S. Daukantas, and A. Lukosevicius, “A comparison of conductive textile-based and silver/silver chloride gel electrodes in exercise electrocardiogram recordings,” *Journal of electrocardiology*, vol. 44, no. 2, pp. 189–194, 2011.

[264] S.-B. Park, Y.-S. Noh, S.-J. Park, and H.-R. Yoon, “An improved algorithm for respiration signal extraction from electrocardiogram measured by conductive textile electrodes using instantaneous frequency estimation,” *Medical & biological engineering & computing*, vol. 46, no. 2, pp. 147–158, 2008.

[265] X. Bao, Y. Deng, N. Gall, and E. N. Kamavuako, “Analysis of ecg and pcg time delay around auscultation sites.,” in *BIOSIGNALS*, pp. 206–213, 2020.

Appendix A

PhyioNet Challenge 2022 Certificate



COMPUTING IN CARDIOLOGY

PhysioNet

Challenge 2022:
Heart Murmur Detection from Phonocardiogram Recordings

Computing in Cardiology and PhysioNet, in recognition of outstanding achievement, present to

Yujia Xu, Xinqi Bao, Hak-Keung Lam and Ernest Nlandu Kamavuako

Team: 'HearTech+'

Challenge 1st Place, Murmur Detection

this certificate of merit for participation with distinction in the 23rd annual Computing in Cardiology/PhysioNet Challenge

Tampere, Finland September 7, 2022

Fig. A.1 PhyioNet Challenge 2022 Certificate