ENABLING ACCURATE CARDIOPULMONARY MONITORING USING MACHINE LEARNING AND A CHEST-WORN WEARABLE PATCH

A Dissertation
Presented to
The Academic Faculty

by

Michael Chan

In Partial Fulfillment
of the Requirements for the Degree
Machine learning in the
Biomedical Engineering Department

Georgia Institute of Technology
August 2023

ENABLING ACCURATE CARDIOPULMONARY MONITORING USING MACHINE LEARNING AND A CHEST-WORN WEARABLE PATCH

Approved by:

Dr. Omer T. Inan, Advisor
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Mozziyar Etemadi
Department of Biomedical Engineering
Northwestern University

Dr. Rishi Kamaleswaran
Department of Biomedical Informatics
Emory University

Dr. Mark A. Davenport
School of Electrical & Computer Engineering
Georgia Institute of Technology

Dr. Thomas Plötz
School of Interactive Computing
Georgia Institute of Technology

Date Approved: [May 17th 2023]

Study hard what interests you the most in the most undisciplined, irrelevant, and original manner possible.

Richard Feynman

To my parents, family, friends, Taiwan, and to everyone who dreams about saving people's lives

ACKNOWLEDGEMENTS

My pursuit of this Ph.D. at Georgia Tech has been a really rewarding one. During my time here, I gave it everything I had. From collecting data, curating data, reading the amazing works the pioneers have done in literatures, designing the algorithms I am proud of, to perfecting the organization of my codebase with each new project, transitioning from hating to loving writing and realizing how much it helped me organize my thoughts, and discussing the cool research ideas we each had with my lab mates, all of these experiences made me feel being in the most utopian research lab. I have been waking up and occupied by my research problems and falling asleep with them being my last thoughts in mind constantly. I really enjoy being a scientist.

First and more most, I would like to thank my family, especially my parents, my sisters, and my grandmoms, for everything they gave me, including the heart to help others, the courage to confront any challenges proactively in life, and the curiosity that eventually led to my pursuit of this Ph.D. degree. They are the reasons why I have been obsessed with building biomedical tools to improve people's health. Thank you, I really miss you and I love you.

I would also like to thank my advisor, Omer, for giving me everything I am looking for in a Ph.D. journey—an opportunity to work on challenging yet very interesting research problems, with many intellectual, inspiring scientists who are scientifically curious, in an environment that organically fosters a collaborative culture. I missed the thrilling weekly CDC research meetings we had since I had to use everything I knew to achieve the best results. Every week, I also tried to bring some new knowledge I learned about the signals,

the physiology, and the diseases to the meetings so we can pretend we were physiological philosophers who were just trying to understand our body better rather than engineers who were driven only by the deadlines. Working with you for the past few years have been very enjoyable and enlightening because of your recognition toward the values of my works and the critical questions you have asked. For that, I am truly grateful. I am looking forward to working with you on more biomedical problems in the next few years.

I would also like to express my gratitude toward all of my committee members. Thank you, Dr. Rishi Kamaleswaran, for pointing me to the coolest, most advanced deep learning techniques and the biomedical problems that are important, as well as providing valuable feedback on my research works. Thank you, Dr. Thomas Plötz, for all of your insightful and philosophical questions. Thank you for asking me what type of ML scientist I want to be known as. The theme of this dissertation was highly motivated by this question. I have also learned a ton from you indirectly through your research papers [1]–[3]. Thank you, Dr. Mozziyar Etemadi, for your pioneering work on the wearable patch used throughout this dissertation and for your vision on its impact for heart failure patients. I am really excited about its progress in the next few years. Thank you, Dr. Mark Davenport, for your time and interest in my dissertation. I can't wait to share with you how we can combine digital signal processing and deep learning together.

I would also like to thank all IRLers. I am inspired by your passion, dedication, creativity, and kindness. Venu, Samer, and Asim were the first people I met when I joined the lab, and I knew immediately that they are the people I want to work with because they work on their research problems as if there is no tomorrow. Thank you, Venu, for the numerous research meetings together and the insightful feedback you shared, for the data

collection you helped out, for reviewing my manuscripts so quickly, and for being a really encouraging friends. I really missed the discussions we had about PPG and physiology. Thank you, Asim, for being such a wholesome human being. You have been working on what the stress research field needs instead of just what is publishable, and that is the most noble thing I have ever seen. I also love bringing my research problems to you because I know you can always see things from an angle that I don't know before. I really believe you will become a well-respected, amazing professor like Omer eventually. Meeting and befriending you is one of the best perks for joining this lab in my opinion. Thank you, Samer, for asking these insightful questions and seeing the values of my work. Thank you Moamen and Burak, for all the data collection studies and research meetings we have done together. Your feedback on my works have been very helpful and have been the keys to major contributions of this dissertation. Thank you, Mohammad and David, I learned a little more about SCG every time I talked to you guys. I am also amazed and motivated by the algorithms you have developed for SCG denoising. Thank you, Göktuğ, for being such a good research partner. I still remember the day you told me about this new *Nature* paper that estimated energy expenditure using IMU sensors placed at shank and thigh and suggested we should also add the knee braces to our study. We both knew we had to do it because we can accomplish what no one has done before, and finding someone who shared a common vision was one of the highlights of this journey. I am really looking forward to working with you more closely in the near future! Thank you, Cem, for being such an important member of the lab already. I wished we had more opportunities to collaborate because of how deeply you think. Thank you, Antonio, for being such a positive, encouraging person in the lab. I also hope I can collaborate with you because I know you

put a lot of thoughts into everything you are building. Thank you, Calvin, for being a critical contributor of almost all of my research projects. Thank you, Afra, for your attitude toward research. Somehow, it has made me feel very energizing after talking to you. Can't wait to see the cool papers you will publish in the future. Thank you, Alex Heller and Andrew Carek, for this great wearable patch you have designed and built. You gave me a playground to explore the intersections between physiology and machine learning without constraints. Thank you, Mobashir. Your pioneering algorithmic works and vision on this wearable patch are the main catalyst of this dissertation. Thank you, Kristy, Tamara, John, Chris, and Quentin, for your help on data collection. Thank you, Sungtae An, Syed Khairul Bashar, Sevda Gharehbaghi, Hewon Jung, Jacob Kimball, Hyeon Ki Jeong, Beren Semiz, Luis Rosa, Oludotun Ode, Nordine Sebkhi, Arpan Bhavsar, Keaton Scherpereel, Michael Cho, Gabby Cester, Trask Crane, Prabhkirat Bindra, Jacob Cook, and Farhan Rahman. I wish I can have a chance to talk to each of you and learn more about you.

I would also like to thank Joyce Tung, my girlfriend, for all the thoughtful acts you have done for me. You helped polish my personalities, my massiness, and my scientific figures (I would work on the big picture sketch, and she would help me with the details). I really enjoy your presence, and I am lucky to experience my Ph.D. journey and your MBA journey with you.

I would also like to thank the friends I have made during my time in Taiwan, the Bay Area, LA, Baltimore, and Atlanta. I left every single one of you to chase after a dream I had, and I am excited to tell you that this dream has become more and more vivid. Eric, Kuo-Wei, and Sray, I also consider meeting you and becoming friends with you one of the luckiest things that have happened to me. Those hour-long conversations about politics,

psychology, being a better self, and building a better future for Taiwan were incredibly stimulating and energizing.

Thank you, Dr. Yi-Kai Lo who is now the CEO of ANEUVO, a neuromodulation startup with a mission to treat spinal cord injury, for guiding my undergraduate research projects when I was at UCLA. I still remember the summer seven years ago when I built my first wearable system with you. Just like my Ph.D. journey, my thoughts were completely occupied by that device. You taught me how to critically assess my own work, which I now consider my strongest skill. Besides the quote that I shared in the third page, Richard Feynman also said, “The first principle is that you must not fool yourself and you are the easiest person to fool.” Pursuing this Ph.D. has helped me become person I have always wanted to be: one that is honest with himself.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

ECG electrocardiogram

SCG seismocardiogram

ACC accelerometer

PPG photoplethysmogram

EDR ECG-derived respiration

SDR SCG-derived respiration

PDR PPG-derived respiration

SQI signal quality index

RQI respiratory quality index

SNR signal-to-noise ratio

PD photodiode

LED light-emitting diode

HR heart rate

SV stroke volume

RR respiratory rate

TV tidal volume

SpO_2 peripheral oxygen saturation

VO_2 oxygen uptake

EE energy expenditure

DSP digital signal processing

ST spectro-temporal

FT Fourier Transform

ML machine learning

DL deep learning

RMSE root mean square error

MAE mean absolute error

MSE mean square error

MAPE mean absolute percentage error

PCC Pearson's correlation coefficient

$\mathbf{R^2}$ coefficient of determination

LOSO leave-one-subject-out

CV computer vision

NLP natural language processing

ASR automatic speech recognition

ADL activity of daily living

SUMMARY

The development of non-invasive instruments to monitor health parameters such as HR, RR, SpO_2, VO_2, blood pressure, body temperature, etc. have had a significant impact in the history of medicine. Without much pain and infection risk, doctors may gather information about a patient's body and make better informed make informed decisions to treat the patient. The emergence of these instruments ever since the 19^{th} and 20^{th} century have already saved many lives and revealed many mysteries of our body. Advancements in miniaturization and digitization technologies in recent decades have further encouraged public's interests to measure health parameters anywhere and anytime, most notably in the form factor of a wrist-worn watch. However, this convenient form factor also presents an inevitable constraint—it cannot capture cardiac vibrations continuously and hence would result in a missed opportunity to examine cardiac function more comprehensively. In contrast, another convenient form factor of a chest-worn wearable patch is a more fitting option.

The chest-worn wearable patch has been designed to capture several physiological signals such as ECG, SCG, and PPG simultaneously. However, since this measurement site is not well-perfused and less studied, it remains uncertain whether it is accurate enough for estimating health parameters. Accordingly, in this dissertation, we have developed a series of algorithms to improve and validate the accuracy of the health parameters estimated. To estimate these health parameters, DSP pipelines that consider the sensing principles and the physiology have been employed for denoising, demodulation, and feature extraction while ML models were used capture the complex relationship between

the extracted features and the target variables. Conventionally, physiological knowledge was used with DSP pipelines to obtain signal representation that could be easily used with ML models sequentially. More recently, advanced ML models such as those based on DL architectures have emerged and outperformed the conventional methods owning to their expressivity to learn more salient, scalable representations of the signals, together with regression/classification simultaneously and synergistically. Nevertheless, we have noticed that these DL methods were also less interpretable and analytical and hence may not deepen our understanding of the physiological signals, the physiology, and the diseases while they become more powerful and advanced. Note that the term interpretability is referred to the ability to interpret a ML model using engineering concepts rather than to interpret it using clinical terms. To avoid this "hidden trap" of ML, we have reconsidered and tested different ways to combine DSP, physiological knowledge, and ML.

Such observation has motivated the three aims presented in this dissertation, each has followed an implicit approach to combine DSP, physiological knowledge, and ML for the estimation tasks. In the first aim, we have constructed the classical DSP pipelines for deriving health parameters and replaced one functional module with ML models for SpO_2 estimation and RR estimation. In the second aim, we have replaced multiple functional modules with ML models for RR estimation and HR estimation. In the last aim, we have designed a DL architecture with DSP-inspired functional modules. Collectively, we have developed algorithms that harness the interpretability of DSP, leverage the flexibility, expressive power of DL, and exploit the data available at hand and the physiological knowledge through MTL. Overall, this work has addressed the accuracy problem of the

chest-worn wearable patch by proposing a new algorithmic direction for merging DSP, physiological knowledge, and ML in a cohesive manner.

CHAPTER 1. INTRODUCTION AND BACKGROUND

1.1 Background

During the Coronavirus disease (COVID-19) pandemic, it has been found that respiratory diseases can be better monitored by combining existing molecular diagnostic techniques with wearables to reduce cost and time [4]. Due to our body's sensitive adaptation to cardiovascular and respiratory abnormalities, these parameters are generally indicative of cardiovascular and respiratory health. This may be explained by animals' dependence on oxygen—it extends beyond mere functionality and is indeed crucial for survival [5].

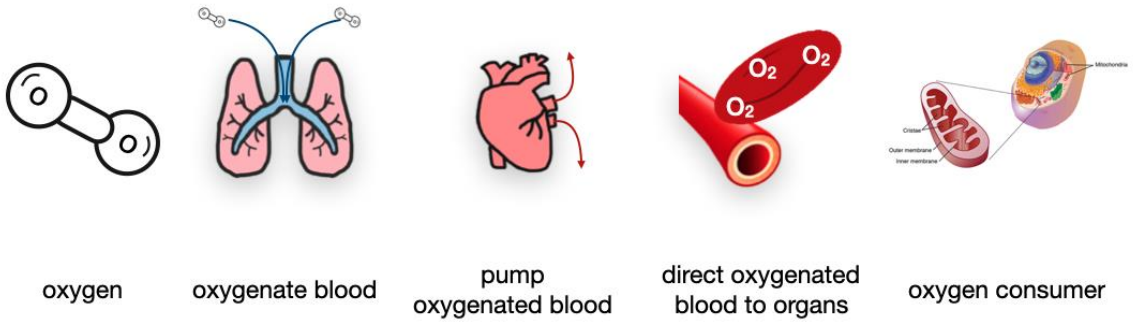


Figure 1.1. Abnormality of the heart and the lung signals underlying pathological conditions.

Oxygen is the most crucial chemicals to transport in the blood, and retraining it at life-sustaining level require coordination of several organs as shown in Figure 1.1. Surprisingly, the oxygen reserve in the blood supports approximately only 3 minutes of metabolic activity at resting state. Further, conscience can be lost within 10 seconds of circulatory arrest [5]. Due to body's low oxygen stores and high demand for oxygen, organs adapt rapidly to maintain oxygen homeostasis. When alveoli are infected with viruses such as COVID-19, the gas exchange efficiency is reduced, which can consequently lead to a

fall in Pa_{O2} and a rise in Pa_{CO2} (and H+ concentration). Due to reduction in arterial oxygenation, lower SpO_2 can be observed. In response, the respiratory centers elevate RR quickly (p. 61) [5]. Elevation of resting HR can be attributed to fever or inflammation [6], but it may also be partially related to hypoxia caused by respiratory dysfunction [7]. Higher VO_2 has been suggested to be indicative of severe COVID-19 infection, while it is also likely a systematic inflammatory response [8]. In the rest of the chapters, we will present methods to estimate these cardiorespiratory accurately.

These cardiorespiratory parameters may be estimated from wearables that are less obtrusive to ADLs. At the same time, these wearables usually offer attractive features such as continuous monitoring, remote monitoring, and potability. The increased popularity and acceptance of wearables may even advance our knowledge in novel disease quickly and comprehensively.

1.2 Motivation

In the year of 2020, Natarajan *et al.* [9] presented a study (*N*=2745) that uncovered the trajectories of health parameters of COVID-19 from patients who were wearing Fitbit watches (Fitbit Inc, San Francisco, CA, USA) within a few months since the outbreak of COVID-19. The study reported that elevated HR and RR of these patients have been observed a few days prior to symptom onset. In the past, such findings were unlikely to be reported so quickly since patients generally would not take vital sign measurements unless symptoms have been present. It is too inconvenient and costly to monitor the people's vital signs continuously without causes. The health wearables have demonstrated their ability to

study disease in an unprecedented way, which may validate our understanding about a novel disease quickly and objectively.

The behavioral changes, interestingly caused by COVID-19 indirectly due to the physical distancing mandates instead of the virus itself, have also been observed using wearable technologies. In a study (*N*=5436) reported in 2021, Capodilupo [10] *et al.* found that WHOOP strap (Whoop, Inc., Boston, MA, USA) users have slept more and exercised more intensely during physical distancing. Such scientific findings would generally be collected from subjective, noisy survey data in the past. Albeit the causal relationship has not been clarified, the correlation between physical distancing and positive behavioral changes have been shown quantitatively. Because of these findings, the society will be encouraged to reflect on how to alter professional commitments and adopt healthier behaviors.

Both cases share similar characteristics in their significance: 1) data collected conveniently from a large pool of users who were already wearing their respective health wearables and 2) reported population-level scientific (pathophysiological or behavioral) findings that were difficult to be observed before. In order to harness the potential of health wearables at the population-level, such devices may need to be fist adopted widely.

1.3 The Accuracy Problem for Health Wearables

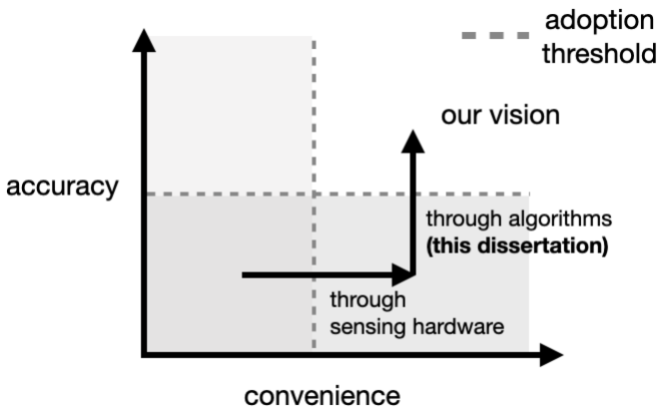


Figure 1.2. The convenience problem (arrow pointing rightward) and the accuracy problem (arrow pointing upward) for health wearables toward public adoption.

For health wearables to become tools for studying disease at the population-level, encouraging behavioral change for healthier life, and informing better treatment plan, they need to become more widely adopted. The overarching question of how to drive public adoption of health wearables can be decomposed into two sub-problems: the convenience problem and the accuracy problem (see Figure 1.2). The convenience problem is what separates the medical instruments used in clinic from health wearables. The main reason why standard medical instruments are not used to extensively monitor health parameters of patients out of clinic is because the measurement devices are expensive, obtrusive to ADLs, pose stigmatization risk, etc. In brief, they are too inconvenient. Luckily, the advancements in miniaturization and digitization technologies in the past few decades have addressed the convenience problem for some health monitoring applications well. The sensing hardware of health wearables have become smaller, cheaper, and more desirable to wear. Despite being more convenient, health wearables may still not be widely adopted due to the accuracy problem. The accuracy problem arises from the unresolved questions of how to improve the true and perceived accuracy of the health wearables, so they can

become more usable. True accuracy describes the estimation accuracy of the health wearables when compared to ground truth devices, and we will address this by improving the estimation algorithms. Meanwhile, the perceived accuracy describes whether patients feel the data is trustworthy and accurate, and poor perceived accuracy could hinder the adoption of the health wearables [11]. In a survey conducted by Li et al. [12] on population that was 60 years old or above, a population that could benefit more from health wearables when compared to other age groups, 34.9% were concerned about the accuracy of health wearables. We will address perceived accuracy by integrating and validating quality assessment methods together with our estimation algorithms. Next, we will explain what construct the wearables physiological signals, study what make the estimations of certain health parameters inaccurate, and propose and evaluate algorithmic approaches to address various aspects of the accuracy problem.

1.4 Non-Invasive Physiological Signals from a Chest-Worn Wearable Patch

The non-invasive physiological signals measured from the chest-worn wearable patch include ECG, SCG, and PPG. Because of the differences in their sensing principles, they each carry different information about our physiological state. The SCG captures the mechanical activity of the heart while the ECG captures its electrical activity. PPG, on the other hand, measures the pulsatile changes in blood volume in arteries, veins, and capillaries. Together, these signal carry hidden information about our cardiovascular system, respiratory system, autonomic nervous system, and their interactions.

In ECG, the amplitude can be modulated by respiration due to the rotation of the cardiac vector [13]. Respiratory sinus arrythmia can modulate HR—derived from the RR

intervals of the ECG—through respiratory centers in the medulla and pons [14]. In SCG, the strength of its amplitude has shown to be informative of SV [15], [16] and can be considered as a surrogate measure of the contractility of the heart. Since the amplitude of contractility of the heart can be modulated by respiration, SCG is also amplitude modulated by respiration [17]–[19]. Chest PPG signal is formed by two major components: cardiac pulsations and respiratory pulsations. The cardiac pulsations represent the changes in arterial blood volume [20], and the normalized amplitude of the arterial PPG in red and infrared can be used to estimate SpO_2 [21]. Respiration can modulate arterial blood and venous blood since the changes in intrathoracic pressure affect blood flow [20]. The vasoconstrictive effect of the sympathetic nervous system on the arteries can also redistribute the blood toward the veins [22]. The respiratory pulsations in PPG may also originate from respiratory movement artifacts because changes in the distance between LED and PD can affect PPG intensities [21]. In **Chapter 2** and **Chapter 5**, we primarily used the cardiac information from these signals for health parameter estimation. In **Chapter 3** and **Chapter 4**, the respiratory information was primarily extracted. In **Chapter 6**, the cardiac and the respiratory information were both used.

1.5 Improve Accuracy with Physiological Knowledge, Digital Signal Processing, and Machine Learning

For a long time, the approaches to improve the accuracy of the health wearables were based on DSP, new physiological and pathophysiological knowledge, and new sensing hardware. As we learned more about the physiological signals measured (e.g., rotation of the electrical dipole caused by respiration [13], [23]), DSP approaches that exploited such knowledge were also developed soon after (e.g., adaptive filter for deriving respiratory

signal from ECG [24]). Because of the new DSP approaches, more hidden information may reveal new knowledge about the pathophysiology (e.g., respiratory variability derived from ECG uncovers correlates of brain activity and vagal nerve stimulation with stress condition [25]). The advancements in physiology research and DSP techniques have been co-evolving together, each influencing and informing the other to drive progress synergistically. In the context of health wearable sensing, such mutualism relationship between the two disciplines have been particularly constructive for addressing the accuracy problem. However, other approaches that can address the accuracy problem have also emerged more recently.

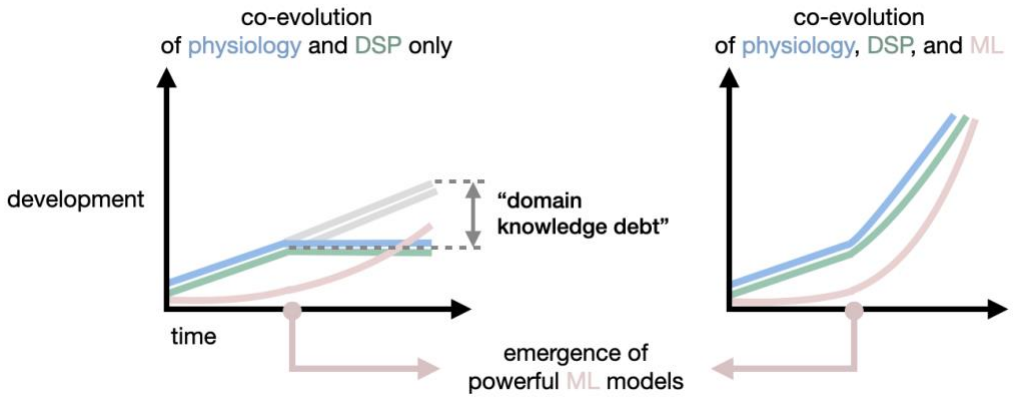


Figure 1.3. Developing methods that consider physiology, digital signal processing (DSP), and machine learning (ML) together is the key to revitalize the co-evolving interactions. (a) Status quo. (b) Vision of this work.

The emergence of sophisticated ML models brought exciting improvements in several fields including health parameter estimation using wearables. By applying black box models that worked in NLP, CV, or ASR to the biomedical problems of interest, we can quickly improve estimation performance when compared to approaches that were based on DSP and physiological knowledge. However, there exists a hidden trap that is

usually overlooked in these quick wins—it can slow down the development of DSP and physiological knowledge. The use of black box models can surely improve the performance of the target task, but the research community is not becoming smarter with the data (sometimes worse, as black box models can be trusted blindly [26]). ML does not seem to have a strong mutualism relationship with DSP and physiological knowledge. Further, as researchers shift their attention toward black box models that do not rely on DSP and physiological knowledge, the progress of the co-evolving DSP and physiological knowledge may even slow down—namely, the "domain knowledge debt," as illustrated in Figure 1.3. Acknowledging this abstract, yet real issue, we have therefore endeavored to reconsidered new ways to marry domain knowledge in DSP and physiological knowledge, and ML in a logical and cohesive manner. Their integration was progressively accomplished through the three aims outlined in this dissertation:

Aim 1: Improve Wearable Accuracy by Replacing One Functional Module in a DSP Pipeline with ML (**CHAPTER 2** and **CHAPTER 3**)

Aim 2: Improve Wearable Accuracy by Replacing Multiple Functional Modules in a DSP Pipeline with ML (**CHAPTER 4** and **CHAPTER 5**)

Aim 3: Improve Wearable Accuracy by Merging DSP and Physiological Knowledge into a DL Model (**CHAPTER 6**)

As the ML models employed became more integrated with DSP and physiological knowledge, we will show that they also became more interpretable. Throughout this dissertation, the term interpretability is loosely defined as the ability to interpret a ML

model using engineering concepts. It is not necessarily related to interpretability in the clinical sense.

To test the performance of the ML models, we collected data with wide dynamic range of the cardiorespiratory responses using protocols that induce physiological changes (e.g., exercise and breath-holds). Although the objective of this dissertation is to develop methods to accurately estimate cardiorespiratory parameters of diseased population from wearables, we chose to validate on the young and healthy (Age: 22-30 years, Weight: 38-105 kg, Height: 153-198 cm) population as the first step to demonstrate wearables' capability to capture physiological changes under controlled conditions.

1.6 Thesis Organization

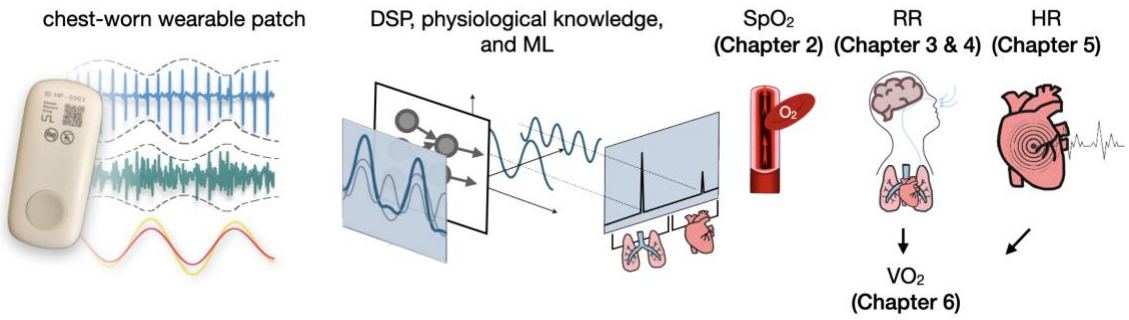


Figure 1.4. The sensing hardware used throughout this work is a chest-worn wearable patch placed at mid-sternum. By combining DSP, physiological knowledge, and ML, we can achieve more accurate estimations of the health parameters.

The objective of the proposed research is to estimate health parameters accurately using DSP, physiological knowledge, ML, and a chest-worn wearable patch as shown in Figure 1.4. The remainder of this dissertation is organized as follows: **CHAPTER 2** describes and evaluates a data-driven method to extract features from PPG for estimating SpO_2. **CHAPTER 3** describes and evaluates a denoising method to enhance features from

ECG and SCG for estimating RR during exercise recovery. **CHAPTER 4** describes and evaluates a modality-attentive, denoising method to fuse ECG and PPG for estimating RR during walking. **CHAPTER 5** describes and evaluates a DL model that extracts features, transforms temporal signals to spectral domain within the DL model, and approximates dominant frequency from SCG for estimating HR. **CHAPTER 6** describes and evaluates a DL architecture designed using DSP and physiological knowledge to estimate VO_2 from SCG, ECG, and PPG. Finally, **CHAPTER 7** provides a conclusion of this dissertation and directions for future work.

CHAPTER 2. ROBUST FEATURE EXTRACTION USING DATA-DRIVEN APPROACH FOR SPO_2 ESTIMATION AT THE STERNUM

In light of the COVID-19 pandemic, SpO_2 has shown to be amongst the vital signs most indicative of deterioration in persons with COVID-19. To allow for the continuous monitoring of SpO_2, we will demonstrate the feasibility to estimate SpO_2 accurately using a custom chest-worn wearable patch, capable of measuring ECG and PPG signals with high fidelity.

The method presented in this chapter was primarily based on the principle of pulse oximetry that is widely used today. Specifically, we improved the procedure to extract ratio-of-ratios of red and infrared PPG, a key feature to relate PPG to SpO_2, by using multiple datapoints in the signal instead of just two. In essence, this pipeline was largely formed using DSP functional modules, and only the feature extraction module was replaced by a data-driven approach proposed. This work demonstrated, for the first time, the feasibility to estimate SpO_2 accurately from the sternum by rigorously validating using a dataset with a large dynamic range, a sufficient sample size (N=20), and an appropriate testing scheme.

2.1 Introduction

Due to the novel COVID-19 pandemic, there is a clear need to monitor respiratory functions in outpatient settings to help assess the progression of COVID-19 during the presymptomatic, symptomatic, and recovery stages. In a recent effort to record and model the trajectories of several vital signs in hospitalized COVID-19 patients, Pimentel *et al.* showed that SpO_2 is amongst the most indicative of parameters of COVID-19 progression

prior to primary outcomes, suggesting the importance to monitor SpO_2 continuously [27]. Through remote SpO_2 monitoring, accurate tracking of COVID-19 progression allows for the implementation of disease-management strategies for both timely interventions and the optimization of scarce medical resources [28].

Unfortunately, existing SpO_2 measurement devices are inconvenient for monitoring in outpatient settings. Typically, SpO_2 is measured through pulse oximeters placed at peripheral extremities such as the fingers; however, these devices obstruct normal ADLs due to restriction of finger usage. In addition, finger-clip based pulse oximeters are accordingly limited in practice to intermittent or single-point measurements. Recently, commercial wrist-worn devices such as Apple Watch, Fitbit Sense Advanced Smartwatch, and Garmin Vivosmart 4 have been developed that allow for more convenient monitoring and offer continuous SpO_2 measurements. Unfortunately, SpO_2 measured at the wrist is likely more susceptible to motion artifacts when compared to measurement sites closer to the center of mass of the body, such as the forehead and ear during walking, as shown by Longmore *et al.* [29]. In addition, peripheral sites such as the wrist respond to apnea events slower than central sites [30], [31] due to the redistribution of blood flow, oxygen conservation [32], and their distal location to the heart.

Thus, central pulse oximeters offer a promising approach for ambulatory outpatient monitoring and in detecting acute hypoxemia events. Furthermore, chest-based pulse oximeters might be advantageous for their synergistic incorporation with other cardiac monitoring methods—such as electrocardiography and seismocardiography—for a more holistic understanding of cardiac functions [33], [15], [34], [35]. Nevertheless, a few major challenges such limitations to the reflectance design, the presence of respiratory artifacts, and the malperfusion of the sternum pose difficulties to the adoption of chest-based approaches. Despite these challenges, prior evaluations have suggested that a chest-based approach might be feasible [29], [36]–[40].

Existing methods using chest-based devices are not rigorously validated, and thus more work is needed to advance chest-worn pulse oximetry. Specifically, validating chest-based pulse oximetry for continuous monitoring would require a sufficiently large and diverse subject population, a wide dynamic range of SpO_2, a resultant RMSE lower than 3.5% [41], and ideally a form factor that does not interfere with ADLs. Upon investigation, there are some clear gaps in the existing literature in these areas that should be addressed. For example, either insufficient sample size (N = 1) in [36] or narrow dynamic range for SpO_2 (88–99%) in [37] limits the validity of the accuracy of their approaches. Meanwhile, Näslund *et al.* [40] showed a strong agreement between their estimated SpO_2 and arterial oxygen saturation (SaO_2); however, their device was unable to capture the pulsatile component of the PPG signals and therefore might be susceptible to motion artifacts and skin pigmentation [21] (p. 34). Kramer *et al.* [38] achieved accurate SpO_2 estimations (RMSE of 2.9%, N = 13), but their work lacks the key details necessary for replicating their algorithm such as preprocessing and feature extraction. Finally, Vetter *et al.* [39] conducted a study following the International Organization for Standardization 9919 international standard, included a moderate number of subjects (N = 10) and a wide dynamic range of SpO_2 (70–99%), and provided sufficient details of their approach. Nevertheless, the prototype appears to be cumbersome to use as it requires a chest-strap to affix the device for sufficient contact pressure. Additionally, due to the lack of LOSO cross validation for training and testing the model, their method is susceptible to data leakage and therefore the high accuracy attained may not be generalizable. To the best of our knowledge, there exists no known accurate chest-based pulse oximetry approach that has been thoroughly validated within the literature. This work attempts to address these shortcomings and present a chest-based pulse oximeter that can estimate SpO_2 in close agreement with a commercially available, validated finger pulse oximeter.

The aim of this chapter is to demonstrate the feasibility of our small, standalone chest-based wearable patch biosensor. In this work, we collected data from 20 subjects who underwent a breath-hold perturbation to induce hypoxemia and used PPG signals from a custom chest-based biosensor to estimate SpO_2. To estimate SpO_2 robustly, we present a novel algorithm to extract key PPG features to account for poor perfusion at the sternum [42] and involuntary respiratory artifacts [43]. In addition, we developed a PPG_{green}-based outlier rejection algorithm for rejecting red and infrared (IR) PPG beats of lower quality. Finally, we demonstrate the optimal calibration scheme for practical usage of this chest-based pulse oximetry. These contributions pave the way for the continuous monitoring of SpO_2. The complete study protocol and algorithmic advancements necessary to achieve accurate chest-based pulse oximetry are described in detail below.

2.2 Methods

2.2.1 Dataset

The breath-hold study was designed to induce hypoxemia and sufficient changes in SpO_2. This study was conducted under a protocol approved by the Georgia Institute of Technology Institutional Review Board (H21100). A total of 22 (16 males, 6 females) young volunteers were recruited for the breath-hold study and written informed consent was obtained. The number of subjects recruited exceeds that of similar studies [39], [44], [45]. In this dataset, two subjects were excluded for analysis. The data of one subject suffered from poor ECG quality—due to expired ECG electrodes that were inadvertently used. The data of the subject exhibited an abnormal distribution of the extracted features compared to those shown in [21] (p. 51). Specifically, the ratio of ratios (R) systematically deviates more than three standard deviations across all SpO_2 levels. Therefore, for this work, only data of the remaining 20 subjects were used for analysis. Demographic

information of these 20 subjects including age, weight, height, Fitzpatrick skin type, perfusion indices (PI), etc. are summarized in Table 1. Note that the distribution of PI for red and infrared in this dataset falls well below the poor perfusion threshold (0.3%) as defined by the Food and Drug Administration (FDA) [41], suggesting this measurement site is indeed malperfused.

Table 2.1 Demographics and physiological responses during breath-hold of the subjects included in the analysis.

Variable.		Female (N = 6)	Male (N = 14)
Demographics	Age (years)	28.00 (2.00)	26.15 (2.19)
	Fitzpatrick skin type (I–VI)	1.83 (1.17)	2.36 (0.93)
	Weight (kg)	56.35 (11.61)	76.44 (11.97)
	Height (cm)	163.03 (7.46)	178.03 (7.09)
	BMI (kg/m²)	21.02 (2.86)	24.05 (2.84)
Breath-hold response	Baseline SpO_2 (%)	96.50 (0.84)	96.29 (1.27)
	Nadir SpO_2 (%)	88.80 (4.81)	80.52 (8.70)
	Breath-hold duration (s)	44.07 (25.64)	55.99 (15.52)
	Approximate finger SpO_2 delay (s)	24.91 (8.69)	27.42 (12.49)
Perfusion index (PI)	Red PPG (%)	0.05 (0.03)	0.07 (0.04)
	Infrared PPG (%)	0.08 (0.04)	0.10 (0.06)
	Green PPG (%)	0.56 (0.23)	0.60 (0.40)

Note values are presented in mean (standard deviation).

In the breath-hold study, subjects were first asked to shave their chest hair to reduce interference. Subsequently, each subject performed 10 end-expiratory breath-holds while sitting in an upright posture with a one-minute break between breath-holds. One minute was found to be sufficiently long for SpO_2 to return to its baseline level. Subjects were instructed to hold their breath for as long as possible. Throughout the study, subjects wore a nose clip and held the disposable mouthpiece (AFT36 bacteriological filter; Biopac System Inc, Santa Barbara, Calif) between their lips. After the data were collected, important oxygenation/deoxygenation events were manually labeled.

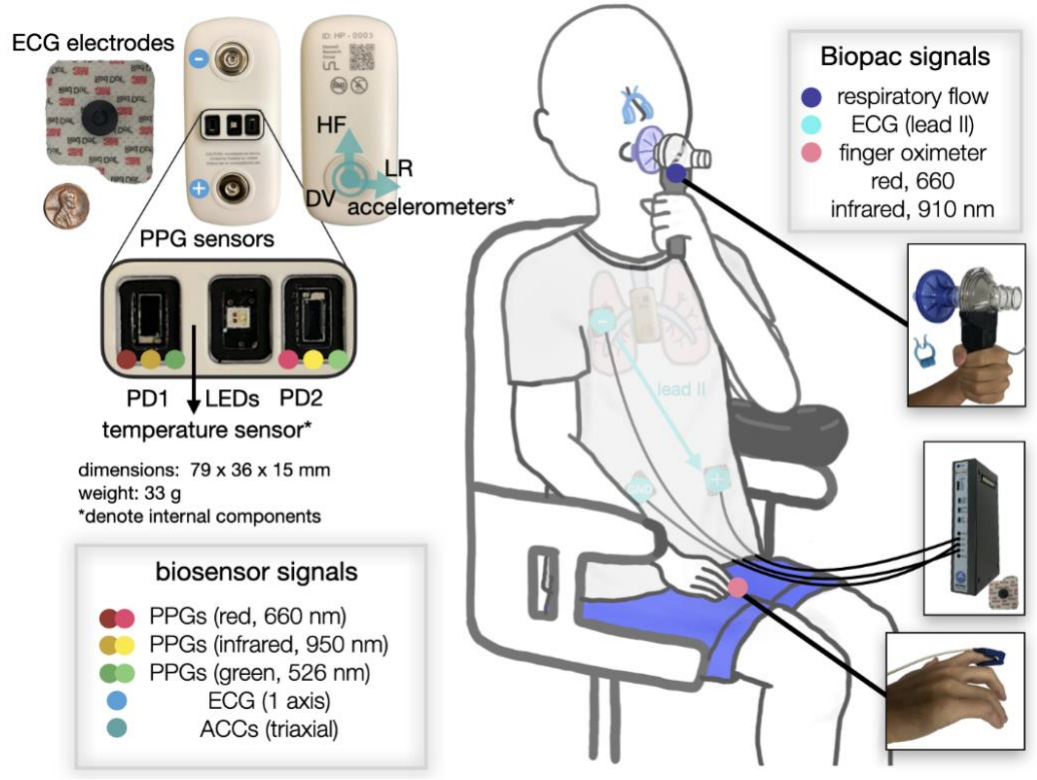


Figure 2.1. Illustration of a subject undergoing the breath-hold study. Placements of the wearable patch biosensor (left side) and the ground truth Biopac sensors (right side) are depicted. The relative size of the biosensor is shown with an off-the-shelf ECG electrode and a penny. Note that the photodiode (PD) has an area of 4.5 mm^2. PD1 is on the left side and PD2 is on the right side of the subject.

As depicted in Figure 2.1, we collected the following information: ECG (Biopac ECG100A; Biopac System Inc, Santa Barbara, CA, USA), right index finger SpO_2 (Biopac OXY100E, TSD124A Finger Clip Transducer; Biopac System Inc), and respiratory flow (TSD117A Medium Flow Pneumotach Transducer; Biopac System Inc) data, all sampled at 2000 Hz. The Biopac OXY100E module reports an accuracy ± 2% for a SpO_2 range of 70–100%. We used the 3M™ Red Dot™ ECG electrodes (model 2660; 3M, Saint Paul, MN, USA) throughout the study. Data outside of this SpO_2 range were discarded since the accuracy is unknown.

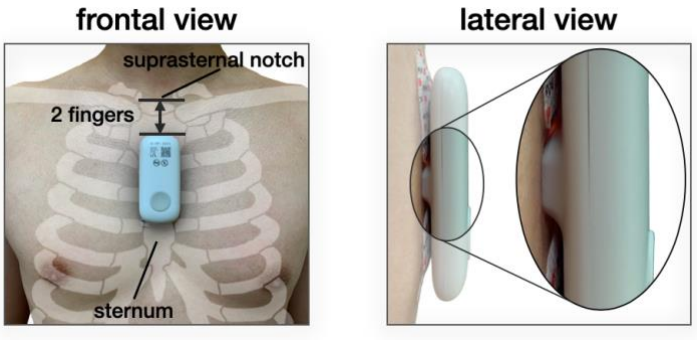


Figure 2.2. From the frontal view, the wearable patch biosensor is attached to the sternum of the subject using ECG electrodes. The superior end of the device starts approximately two fingers down from the suprasternal notch. From the lateral view, the protruded part of the device that houses the LEDs and PDs is visible. Note that light is hardly visible from the sides of the device.

In parallel, we also attached the wearable patch biosensor to the subject's mid-sternum and collected single-lead ECG, two sets of multiwavelength PPGs (red, infrared [IR], and green), and tri-axial SCG (not used in this study), sampled at 500, 67, and 1000 Hz, respectively. The hardware used in the biosensor is almost identical to that reported in our previous work [15], [35], [46] except for the addition of the PPG modules and the change in form factor. The ECG analog front-end (AFE) and the accelerometer AFE (for SCG) remain the same. Specifically, the PPG AFE used to drive the LEDs and obtain data from the PDs is the Maxim 86170 (Maxim Integrated, San Jose, CA, USA). The multi-chip LEDs, which has red (660 nm), and IR (950 nm), and green (526 nm) wavelengths, are the SFH 7016 (OSRAM, Munich, Germany), and the PDs are the VEMD 8080 (Vishay Semiconductors, Heilbronn, Baden-Württemberg, Germany). Serial Peripheral Interface was used as the communication protocol between the microcontroller and peripheral sensors. This device is also equipped with wireless capabilities (i.e., Bluetooth and Wi-Fi) for transmitting data. However, in this study, data were stored in the Secure Digital card and later retrieved by a custom-built software application as in previous work [15], [35],

[46]. The battery life of the device at the full sample rates of all sensors is up to 60 hours. The front and lateral views of the device are shown in Figure 2.2.

2.2.2 Manual Labelling

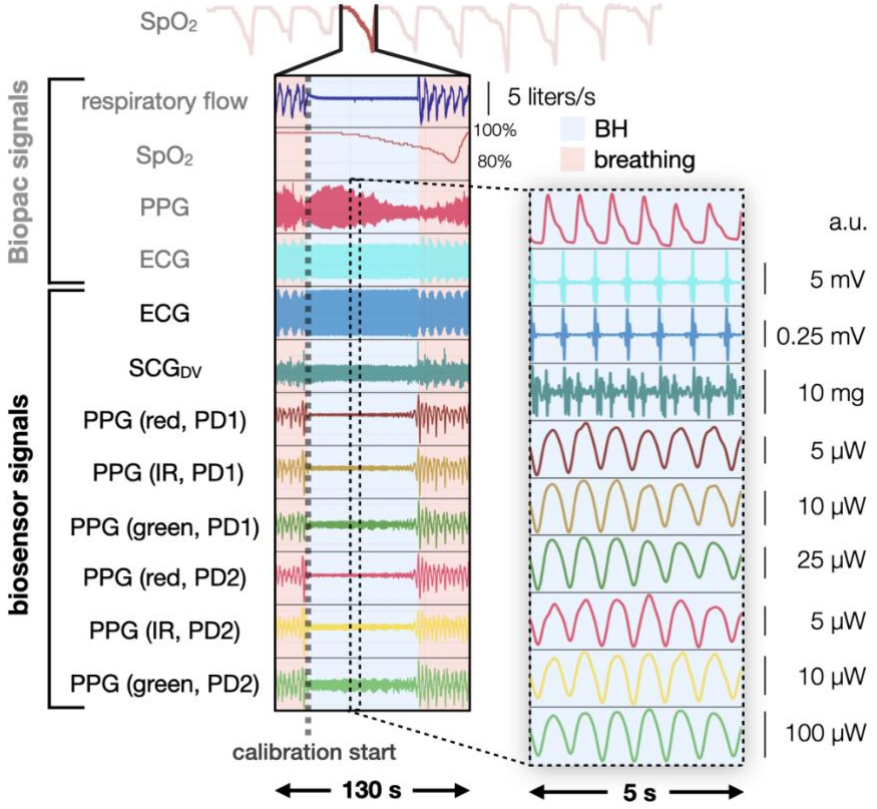


Figure 2.3. Biopac and biosensor signals during the fourth breath-hold (BH, blue) and during normal breathing (pink). A 5s window during breath-hold is shown to illustrate the quality of the signals.

In Figure 2.3, filtered, high-quality physiological signals acquired during breath-hold and breathing are presented. For each subject, we selected the PD with higher quality as determined by visual inspection. The discrepancy can be attributed to the differences in LED/PD separation distance as LED/PD separation distance can affect the quality of PPG [21] (p. 88). Further assessment may be needed since optimizing the LED/PD separation distance is a critical factor for obtaining a good quality signal. Manual labeling was

performed using the respiratory flow and the SpO_2 data. Alignment was necessary since it has been observed that deoxygenation events do not occur simultaneously for different body sites, and SpO_2 measured at the finger is usually delayed from SpO_2 measured at central sites [30], [31], [47]. This delay can be partially attributed to the oxygen-conserving effect induced by breath-hold. Similar to the diving response [32], breath-hold also leads to bradycardia and peripheral vasoconstriction to reduce oxygen consumption in peripheries and redistribute blood flow to vital organs such as the brain and the heart [32]. The combined effect leads to a delayed deoxygenation measurement by a finger-pulse oximeter when compared to a pulse oximeter placed closer to the heart or the brain. Davies *et al.* reported a mean delay of 16.75 ± 5.88 s across subjects for their in-ear reflectance pulse oximetry [30].

From the respiratory flow (top signal in Figure 3), breathing (the ”oscillating” part, pink) and breath-hold (the “silent” part, blue) segments can be easily distinguished. From the ground truth SpO_2 (the second signal on the left in Figure 3), three distinct timestamps were recorded of each deoxygenation event: start, nadir, and end. The start of deoxygenation is defined as the point where SpO_2 begins to drop drastically (rate of SpO_2 decline > 0.5% / cardiac cycle for 3 consecutive cardiac cycle). The nadir of deoxygenation is defined as the lowest SpO_2 within the deoxygenation event. The end of deoxygenation is defined as the point where SpO_2 returns to the baseline level. Usually, the nadir and the end of deoxygenation can be easily identified. To account for the delay of deoxygenation between finger and chest deoxygenation, the nadir deoxygenation of the finger is aligned to the end of the breath-hold, based on the assumption that chest arteries received well-oxygenated blood immediately following the end of the breath-hold. Though our results

suggest this alignment procedure is somewhat accurate, we found that a more precise alignment algorithm was required to achieve adequate accuracy; the updated alignment algorithm is applied and provided in Appendix A.

2.2.3 Principle of PPG

The PPG signal used in this work represents the changes in the reflected LEDs light intensity, as detected by the PDs. According to the Beer–Lambert law, the intensity of the reflectance PPG measured is related to the intensity of the incidence light, the absorption coefficients of the absorbers, and the optical path length of light traveled from the LEDs to the PDs [21] (pp. 47–48). The changes of PPG intensity with respect to each component (arterial blood, venous blood, tissue, bone, etc.) have different pulsating dynamics [48]. By using appropriate filter banks, we can leverage the cardiac pulsation of the PPG to target its arterial, pulsatile, small-signal component. Specifically, the portion of the PPG signal that is representative of cardiac pulsation and the periodic changes in blood volume is termed the alternating current (AC) component, and the baseline wander of the PPG—which is slower than the cardiac frequency—is termed the direct current (DC) component [48]. The AC and DC components of the PPG in multiple wavelengths (i.e., red and IR PPG) can reveal the oxygenation saturation of the underlying arteries [21]. More details are provided in the following sections.

2.2.4 Signal Processing Pipeline

2.2.4.1 Principle of Pulse Oximetry

Principle of To relate the aligned signals to ground truth SpO_2, relevant features in the biosensor signals need to be extracted. We followed the standard approach described in [21] (p .131) and tailored the algorithm to our pulse oximetry. The key feature, R, defined

as the ratio of the normalized AC component (also a ratio) of two optical wavelengths, can be extracted from the PPG signals:

$$R = \frac{\frac{AC_{red}}{AC_{IR}}}{\frac{DC_{red}}{DC_{IR}}} \quad \textbf{(2.1)}$$

where AC_{red} is the AC component of the red PPG, AC_{IR} is the AC component of the IR PPG, DC_{red} is the DC component of the red PPG, and DC_{IR} is the AC component of the IR PPG. Normalization is performed by dividing the AC component of a wavelength by its DC component. R, along with the absorption coefficients of oxyhemoglobin (HbO_2) and deoxyhemoglobin (Hb) for different wavelengths, can be used together to derive SaO_2 directly. According to [21] (p. 50), the theoretical relationship between SaO_2 and R is defined as:

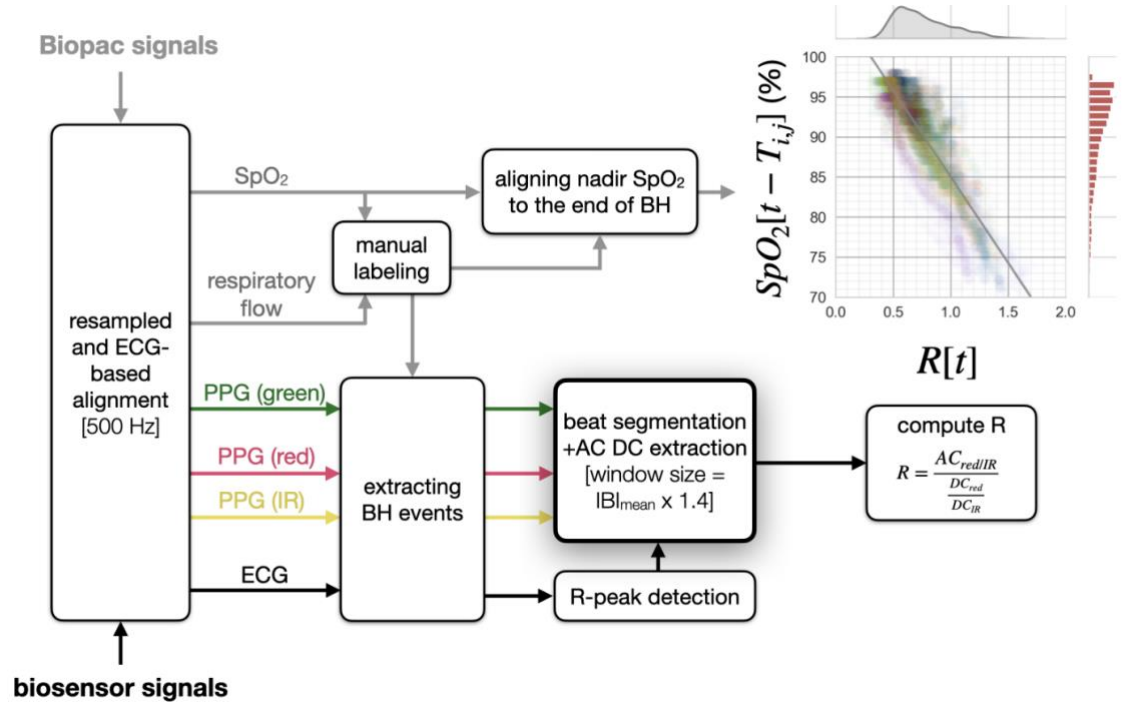
$$SaO_2 = \frac{\varepsilon_{Hb}(\lambda_{red}) - \varepsilon_{Hb}(\lambda_{IR})R}{\varepsilon_{Hb}(\lambda_{red}) - \varepsilon_{HbO_2}(\lambda_{red}) + \left[\varepsilon_{HbO_2}(\lambda_{IR}) - \varepsilon_{Hb}(\lambda_{IR})\right]R} \quad \textbf{(2.2)}$$

where ε_{Hb} is the absorption coefficients of Hb, ε_{HbO2} is the absorption coefficients of HbO_2, λ_{red} is the wavelength of the red PPG, and λ_{IR} is the wavelength of the IR PPG. If further approximated using a Taylor series expansion,

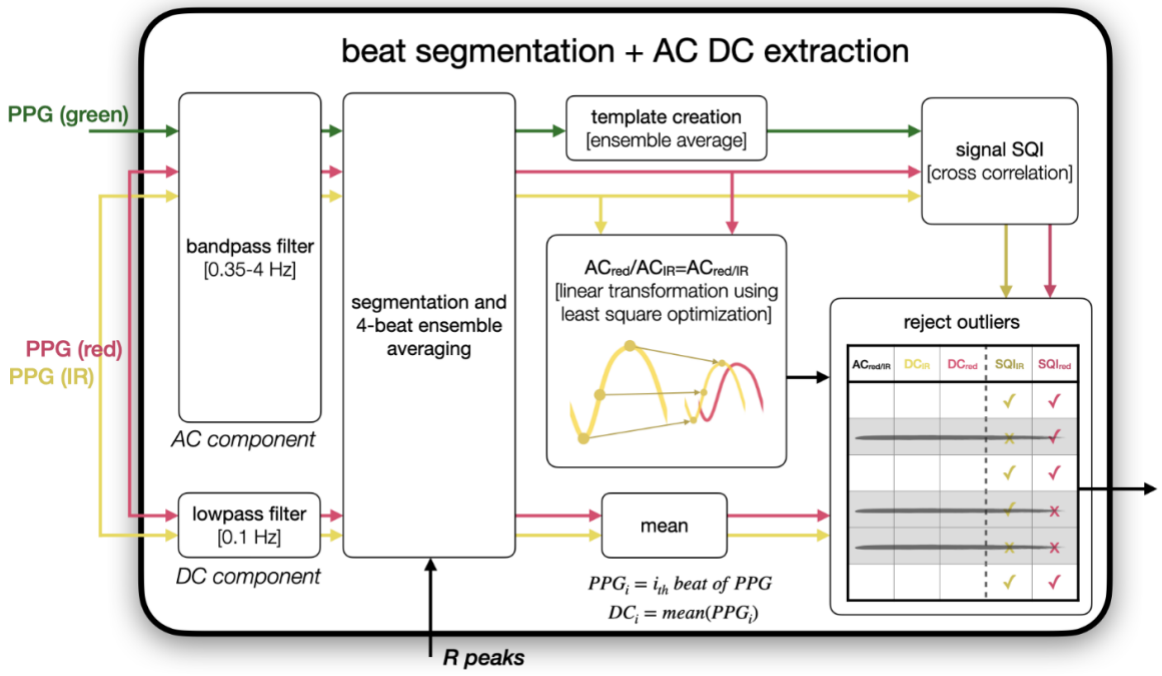
$$\widehat{SaO_2} = \frac{A \times R + B}{C \times R + D} \approx m \times R + b = SpO_2 \quad \textbf{(2.3)}$$

emerges as an empirical model that governs the relationship between SpO_2, the surrogate of SaO_2, and R. In Equation (3), A, B, C, and D replace the absorbance coefficients of the Hb and HbO_2 at the two wavelengths [21] (p. 54), *m* is the slope, and *b* is the intercept.

2.2.4.2 Preprocessing Overview



(a)



(b)

Figure 2.4. Feature extraction block diagrams. (a) Overview of the signal preprocessing and manual labeling pipeline. Extracted pairs of ratio of ratios (R) and SpO_2 were used for training and calibration of the model. $T_{i,j}$ denotes the delay found between R and SpO_2 for the j^{th} breath-hold of subject i. The aligned data were displayed in the upper right scatter plot. The distribution of R and SpO_2 are also shown above and on the right of the scatter plot, respectively. IBI_{mean} denotes the mean of the interbeat intervals (IBI) of a subject. (b) The block diagram for PPG beats segmentation, $AC_{red/IR}$ extraction, DC feature extraction, and PPG_{green}-based outlier rejection.

Figure 2.4 (a) illustrates our signal-preprocessing pipeline, which we used to extract R from the wearable patch biosensor signals. Although there are other signals in the dataset, we only found six that are relevant to this work, namely, finger SpO_2, red PPG, IR PPG, green PPG, ECG, and respiratory flow. The Biopac and biosensor signals were first resampled to 500 Hz and synchronized by maximizing the cross-correlation of their ECG signals. The Biopac SpO_2 was further aligned to the biosensor signals on a per breath-hold basis, using the manual label described in Section 2.3. Due to both respiratory artifacts when emerging from breath-hold and the lack of range in measured SpO_2 values, we only target SpO_2 estimation during the breath-hold segments of the signals. Each extracted R from the breath-hold segments was paired with the manually aligned SpO_2, and both a scatter plot—demonstrating the correlation between R and SpO_2—and distributions plots to show the skewness of R and SpO_2—are depicted in Figure 2.4 (a). The skewness of R and SpO_2 can be partially attributed to our ability to maintain oxygenation homeostasis, enabled by the continuous supply of oxygen by the oxygen stores upon breath-hold [32]. Before uncovering the relationship between R and SpO_2, we will first demonstrate robust feature extraction and the outlier rejection algorithms necessary to extract R for the chest-based pulse oximetry.

2.2.4.3 Learning Features Robustly using Linear Model Parameter

To compute R, it is necessary to compute AC features and DC features of each PPG beat. Note that a feature represents a scalar value to represent the characteristic of a PPG beat in the context of this work. In Figure 2.4 (b), the block diagram for beat segmentation and feature extraction has been shown. To isolate the AC component, an empirically validated bandpass filter, with a passband of 0.35 to 4 Hz was first applied. The low cutoff was chosen to remove the baseline wander, due to involuntary respiratory movement. The high cutoff was chosen empirically so as to reduce the dicrotic notch and preserve only the frequency components with less variation across wavelengths. The AC component was segmented into PPG beats using ECG R-peaks detected by a Pan-Tomkins algorithm [9], modified for R-peak correction and further smoothed using 4-beat ensemble averaging. Conventionally, computing AC features for red and IR PPG relies on robust peak and valley extraction. Although we were able to minimize respiratory artifacts through the breath-hold protocol, involuntary respiratory movements were still present and observable in some subjects. Evidently, extracting R robustly from respiration-corrupted PPG can be challenging [37], [50]. Conventionally, the peak and valley of PPG in each cardiac cycle are extracted to compute the AC features [21] (pp. 129–130). In a preliminary analysis, we found that this method is not reliable as the signal can be easily distorted by the subtle—yet still significant—involuntary respiratory movements at this low perfusion site. To address this, we introduced a novel algorithm that does not require peak and valley extraction. Specifically, by rearranging the terms in Equation (1), we can obtain,

$$R = \frac{\frac{AC_{red}}{AC_{IR}}}{\frac{DC_{red}}{DC_{IR}}} = \frac{AC_{red/IR}}{\frac{DC_{red}}{DC_{IR}}} \quad (2.4)$$

By computing the $AC_{red/IR}$, the ratio of AC_{red} to AC_{IR}, we can avoid the difficulty of extracting peaks and valleys in distorted PPG signals. To do so, we leveraged the fact that the IR PPG beat, denoted as PPG_{IR}, appears to have a similar morphology to the red PPG beat, denoted as PPG_{red}, after being bandpass filtered. Therefore, we can model the relationship of the two PPG beats using a linear transformation method:

$$\alpha_1^\star, \alpha_2^\star = \underset{\alpha_1, \alpha_2 \in \mathbb{R}}{argmin} \|PPG_{red} - \alpha_1 PPG_{IR} - \alpha_2\|_2^2 \quad (2.5)$$

where $PPG_{red}, PPG_{IR} \in \mathbb{R}^N$, N is the number of samples in the PPG beat, and $\alpha_1^\star, \alpha_2^\star$ denote the scale and the bias that will minimize the ℓ^2-norm of their differences. With the assumption that the differences, once optimized, should be closely distributed, we rejected beats with differences of more than five median absolute deviations from the median, which is a more robust rejection criterion compared to the "standard deviations around the mean" method [51]. Note we rejected only 1.79% using this method. The optimal scale, $\alpha_1^\star$, represents the ratio of the AC component of the two wavelengths:

$$AC_{red/IR} = \alpha_1^\star \quad (2.6)$$

In parallel, the DC component was isolated using a low-pass filter with a high cutoff frequency at 0.1 Hz. This cutoff was based on a heuristic assumption that physiological dynamics of faster than 0.1 Hz (e.g., involuntary respiratory movement) do not directly relate to the deoxygenation induced by breath-hold based on data shown in [52]–[54]. The

DC component was similarly segmented and smoothed to ensure consistency with the processing steps for AC extraction. Finally, DC features were computed as the mean of the segmented DC beats.

2.2.4.4 PPG_{green}-Based Outlier Rejection

Although we carefully selected the parameters of the preprocessing and feature extraction pipeline, some PPG beats may still be distorted due to motion artifacts and involuntary respiratory movements and therefore can hinder accurate SpO_2 estimation. Hence, we designed a novel outlier rejection algorithm using the green PPG beats as a signal quality template for its robustness against noise [55], so as to reduce the contamination of abnormal features extracted. Our signal quality assessment relied on two assumptions: (1) reliable red or IR PPG beats in the bandwidth filtered constitute a morphology similar to that of green PPG beats; and (2) outliers in AC ratios are defined as datapoints that deviate by more than five median absolute deviations from the median (similar to the $AC_{red/IR}$ rejection method). To determine the similarity, we consider a methodology described in [56]. First, the normalized cross-correlation (NCC) between a PPG beat with its corresponding template is computed:

$$NCC_{k,\lambda} = \frac{\sum_{n=1}^{N}(PPG_\lambda(n) - \overline{PPG_\lambda})(PPG_{green}(n+k) - \overline{PPG_{green}})}{\sqrt{\sum_{n=1}^{N}(PPG_\lambda(n) - \overline{PPG_\lambda})^2 \sum_{n=1}^{N}(PPG_{green}(n+k) - \overline{PPG_{green}})^2}} \quad (2.7)$$

where $PPG_\lambda(n)$ denotes the n^{th} sample in the PPG beat of wavelength λ, $\overline{PPG_\lambda}$ denotes the average value of the samples in a PPG beat of wavelength λ, and $NCC_{k,\lambda}$ denotes the correlation coefficient between PPG_λ and the k-lag PPG_{green}. Next, the maximal NCC_λ, $NCC_{max,\lambda}$, defined as,

$$NCC_{max,\lambda} = \underset{k}{argmax}\, NCC_{k,\lambda} \tag{2.8}$$

is selected as a measure of the SQI of the PPG_{λ} and has a range of [0, 1]. Both assumptions translate directly to the two upper right blocks in Figure 2.4 (b). Each SQI method has an empirically determined threshold, 0.7, and a sample is excluded if either SQI method suggests so. The PPG_{green}-based outlier rejection algorithm rejected nearly 8.07% of the beats.

2.2.4.5 Computation of R

The output matrix in Figure 5(b) has a dimension of $N_{beats} \times 5$, where the five columns represent the AC features ($AC_{red/IR}$), two DC features (DC_{red}, DC_{IR}), and the two binary SQI decisions (SQI_{red}, SQI_{IR}). Only features approved by the SQI algorithm were used to compute R. Note that we also experimented with the peak and valley method, but it would require a more aggressive outlier rejection threshold (~30% rejection ratio) to achieve comparable SpO_2 estimation accuracy. Finally, R is computed by dividing $AC_{red/IR}$ by $\frac{DC_{red}}{DC_{IR}}$ as shown below,

$$R = {AC_{red/IR}}\Big/{\frac{DC_{red}}{DC_{IR}}} = {\alpha_1^{\star}}\Big/{\frac{DC_{red}}{DC_{IR}}} \tag{2.9}$$

In this dataset, R is a unit-less measure and generally ranges from 0.4 to 1.6 for SpO_2 above 70%.

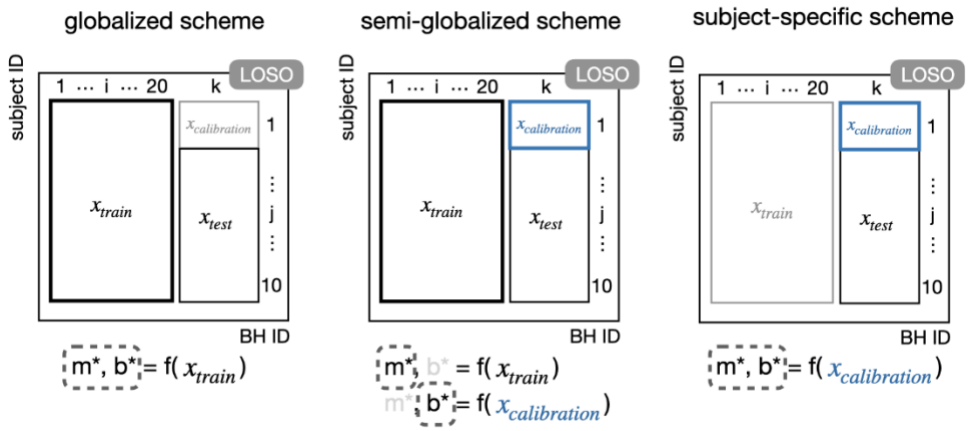


Figure 2.5. Different model training schemes, where x_{train} denotes the data from the training subjects; $x_{calibration}$ denotes the first breath-hold data from the test subject; x_{test} denotes the data of the test subject excluding the first breath-hold; k denotes the left-out subject. The parameters used to construct the model are in the dashed boxes. Data or parameters not used at all were colored in light gray. (a) Globalized scheme (no subject-specific calibration). (b) Semi-globalized scheme (with subject-specific calibration). (c) Subject-specific scheme.

2.2.5 SpO₂ Estimation

2.2.5.1 Linear Regression

The temporally aligned SpO_2 and the extracted R were subsequently used to train the parameters in Equation (**2.3**). The parameters *m* (slope) and *b* (intercept) were estimated by minimizing the ℓ^2-norm of the difference between the ground truth SpO_2 and estimated SpO_2:

$$m^\star, b^\star = \underset{m,b \in \mathbb{R}}{argmin} \|SpO_2 - mR - b\|_2^2 = f(x) \tag{2.10}$$

where *x* denotes pairs of SpO_2 and R, and *f* represents an arbitrary function for determining the optimal parameters of an objective function.

2.2.5.2 Training and Calibration Schemes

Since including a one-time, short calibration procedure is realistic for practical usage of the device, we also investigated the best training and subject-specific calibration procedure. Three training and calibration schemes were considered, including a (1) globalized scheme containing subject-independent training (see Figure 2.5 (a)); (2) semi-globalized scheme featuring global training with subject-specific calibration (see Figure 2.5 (b)); (3) subject-specific scheme (see Figure 2.5 (c)). The globalized scheme is equivalent to the standard LOSO cross validation. The semi-globalized scheme described herein is analogous to the semi-globalized method discussed in [57], aside from the fact that we used duration rather than number of points to standardize the subject-specific calibration. Particularly, in the semi-globalized scheme the globally trained intercept *b* was replaced by a subject-specific calibrated *b* (using the data in the first of the 10 breath-holds). The subject-specific scheme involved training both parameters using only the first breath-hold data of the subject. Note that we also explored calibration using *m*, but the results were considerably worse and therefore not reported. In both globalized and semi-globalized schemes, LOSO cross validation was also used to assess generalizability of the models trained. To compare model performance fairly and to avoid data leakage, we excluded the first breath-hold of the test subjects for evaluation of the globalized schemes to ensure identical testing data.

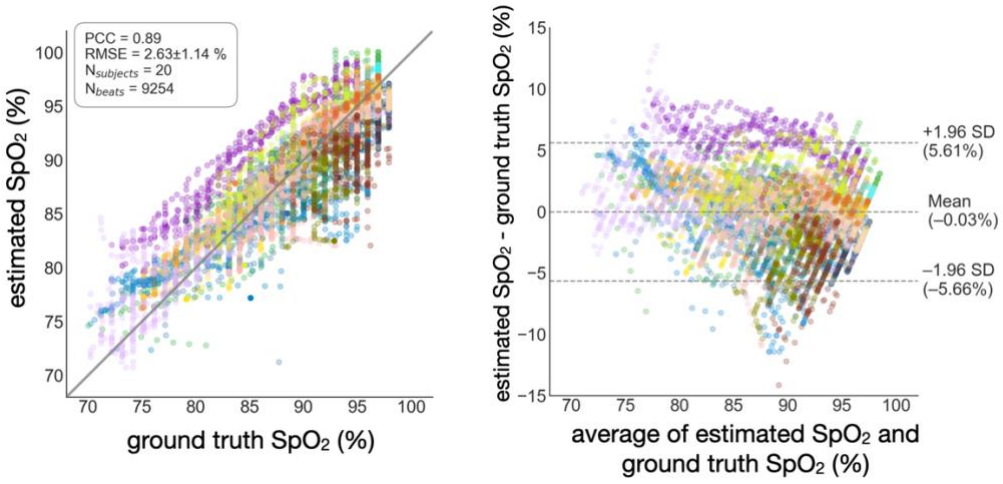
2.2.5.3 Evaluating Model Performance

To assess the performance of these three schemes, we recorded the RMSE, the parameters of the linear model on a per subject basis, and the PCC of estimated SpO_2 on

all subjects jointly. The mean and the standard deviation of the subject-specific RMSEs were computed to summarize the performance of each scheme and subsequently used as the critical metric to assess the capability of the pulse oximetry. Note that the errors presented in this work are all absolute errors rather than relative/percentage errors. The unit of RMSE is denoted by %, which represent the oxygen saturation level.

2.3 Results

2.3.1 Accuracy of SpO_2 Estimation



(a) (b)

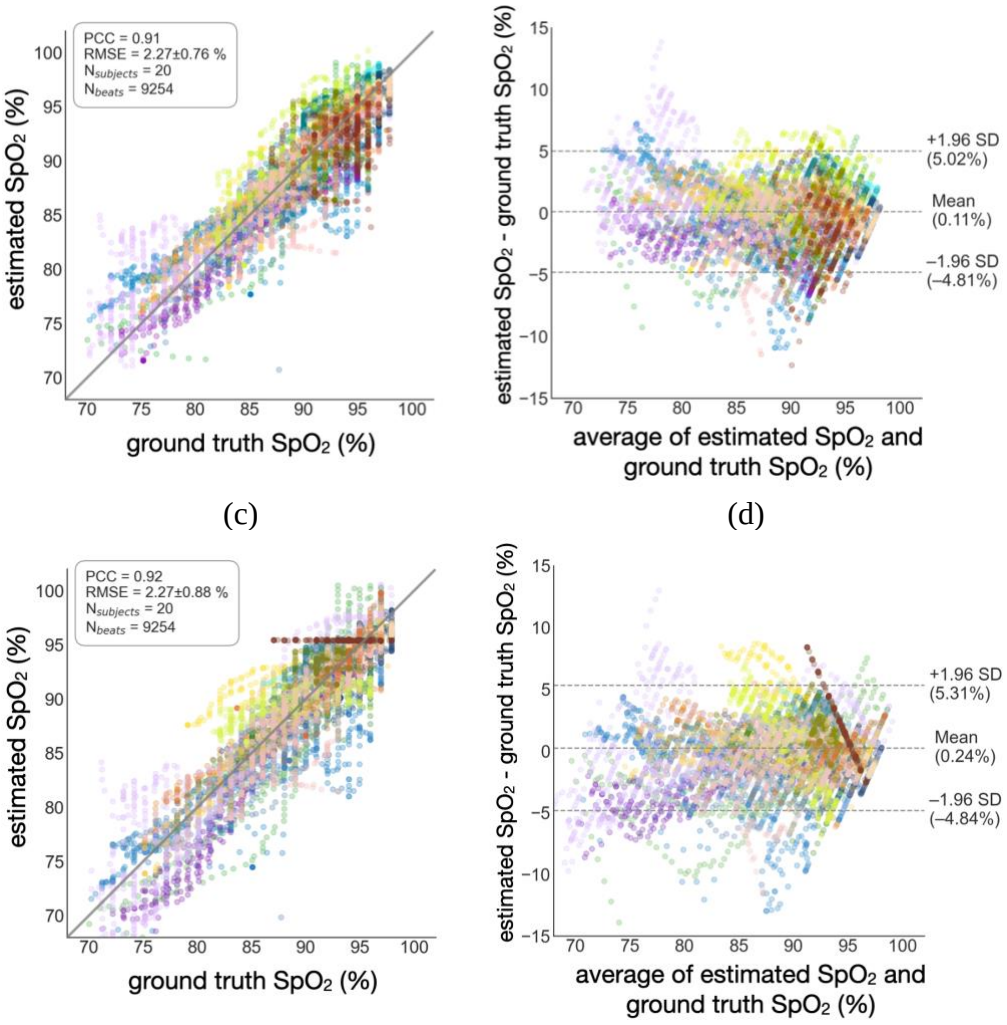


Figure 2.6. Estimation performance of the wearable patch biosensor. The root-mean-square error (RMSE) across subjects and Pearson correlation coefficient (PCC) are shown. Note each color represents a subject. (a) Regression analysis of the globalized scheme. (b) Bland–Altman analysis of the globalized scheme. (c) Regression analysis of the semi-globalized scheme. (d) Bland–Altman analysis of the semi-globalized scheme. (e) Regression analysis of the subject-specific scheme. (f) Bland–Altman analysis of the subject-specific scheme.

Table 2.2 Comparing the results of three training schemes based on their RMSE, PCC, bias, and 95% LOR.

	RMSE	PCC	bias	95% LOR
Globalized	2.63 ± 1.14%	0.89	–0.03%	[–5.66%, 5.61%]
Semi-globalized	2.27 ± 0.76%	0.91	0.11%	[–4.81%, 5.02%]
Subject-specific	2.27 ± 0.88%	0.92	0.24%	[–4.84%, 5.31%]

In Figure 2.6, regression plots and Bland–Altman plots are provided to demonstrate the estimation results. We also summarize the RMSEs across subjects, PCC, bias, and 95% limits of agreement (LOR) in Table 2.2. The globalized scheme achieves lowest accuracy (see Figure 2.6 (a) and (b)). The semi-globalized scheme shows better accuracy (see Figure 6 (c) and (d)). The subject-specific scheme achieves the best accuracy (see Figure 2.6 (e) and (f)). Using the semi-globalized model, we were able to lower the mean RMSE by 0.36% and increase PCC by 0.02 when compared to the globalized model. The semi-globalized scheme and the subject-specific scheme have similar performance levels, both of which are superior to the globalized scheme. From the Bland–Altman plots, both models show minimal bias.

2.3.2 Semi-Globalized Scheme vs. Subject-Specific Scheme

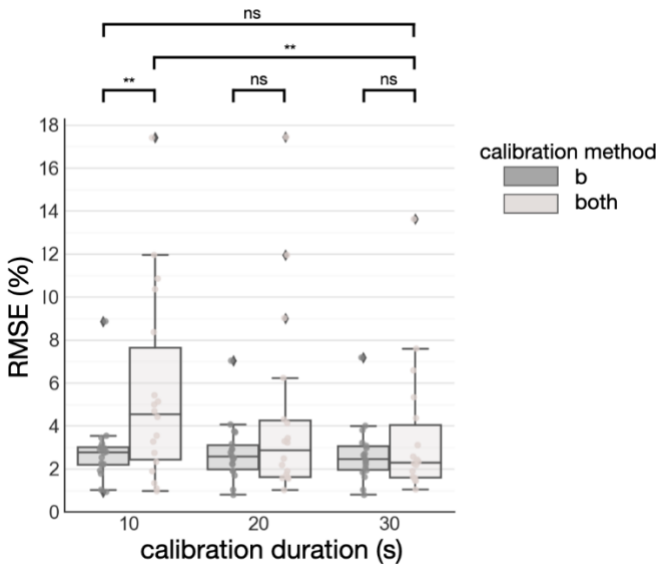
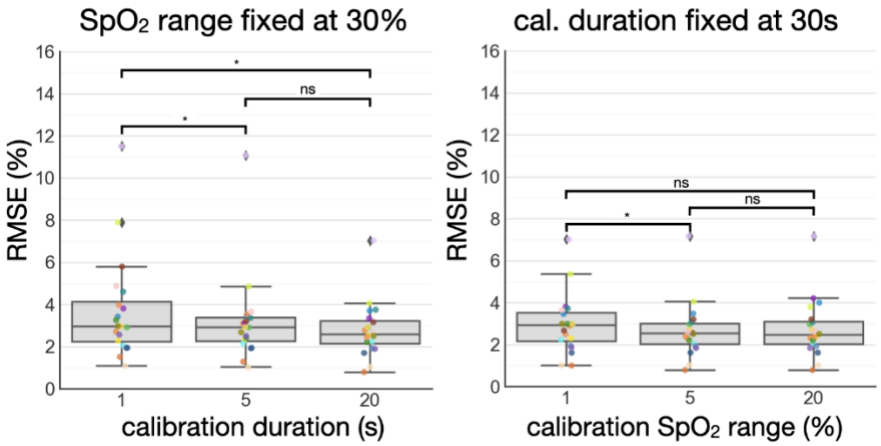


Figure 2.7. Calibration using *b* and using both *b* and *m*. *ns* denotes not significant ($p > 0.05$), and a double asterisk denotes significant differences ($p < 0.01$) as determined by paired sample t-tests.

Since it has not been previously examined in the literature, we also studied which parameters benefit the most from subject-specific calibration. This is accomplished by comparing the semi-globalized scheme (i.e., calibrating b) to the subject-specific scheme (i.e., calibrating both m and b) while varying the calibration duration constraints. The duration constraint was imposed by considering data only within the said duration. Surprisingly, the semi-globalized model works more efficiently at reducing RMSE, as shown in Figure 2.7. Note that two outlier subjects were excluded for better visualization. In all three duration constraints (10s, 20s, and 30s), the semi-globalized schemes achieved a lower RMSE. When comparing the RMSE of calibrating b, constrained by a calibration duration of 10s, to calibrating both parameters by a calibration duration of 30s, we found no statistical significance ($p > 0.05$) as determined by a paired sample t-test. Therefore, we determined that the semi-globalized scheme is the best calibration strategy for this dataset

as it would require a shorter duration to achieve similar performance to the subject-specific scheme.

2.3.3 Standardizing Subject-Specific Calibration: Duration vs. SpO_2 Range.



(a) (b)

Figure 2.8. Calibration analysis. (a) RMSE is reduced significantly with increasing calibration duration while fixing SpO_2 range. (b) RMSE does not change significantly with increasing SpO_2 range while fixing calibration duration. *ns* denotes not significant ($p > 0.05$), and a single asterisk denotes significant differences ($p < 0.05$) as determined by paired sample t-tests.

To study the most efficient way to collect data for calibration, we also examined the changes in RMSE by imposing different constraints on the calibration data, including a duration constraint and SpO_2 range constraint. Similar to the duration constraint, the SpO_2 range constraint considers data only within the said SpO_2 range. According to the results shown in Figure 2.8 (a), we found that increasing calibration duration from 1s to 20s while fixing SpO_2 range to 30% leads to significant ($p < 0.05$) reduction in mean RMSE across the subjects. On the other hand, the results shown in Figure 2.8 (b) suggest that increasing the calibration SpO_2 range from 1% to 20%, while fixing the duration to 30s, did not lead to a significant ($p > 0.05$) difference in mean RMSE. Paired sample t-tests were used for

the statistical analysis. Standardizing calibration duration appears to be the best calibration strategy here.

2.3.4 Effect of Varying Melanin Content

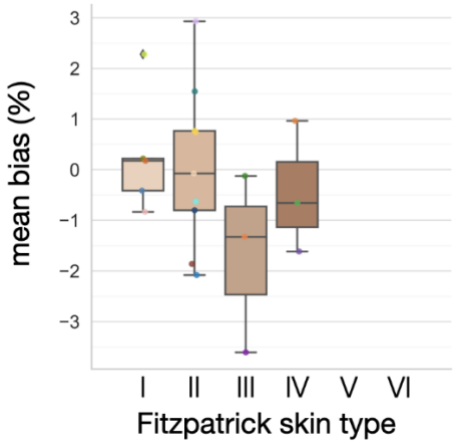


Figure 2.9. Mean bias across varying Fitzpatrick skin types. The differences between Fitzpatrick skin types are not statistically significant ($p > 0.05$). One-way analysis of variance (ANOVA) was used for statistical analysis.

Since none of the subjects had nail polish on their right index finger or tattoos on their sternum, we only considered the confounding effect of the difference in melanin content. In this analysis, we assessed melanin levels using self-reported Fitzpatrick skin types [58] and studied the way melanin content affects the bias between finger and sternum SpO_2. In Figure 9, the errors between different Fitzpatrick skin types are shown to be statistically insignificant ($p > 0.05$), using a one-way analysis of variance (ANOVA). This implies that our device does not introduce different bias for subjects with varying skin melanin content when compared to the finger pulse oximeter, understandably, as both operate on the same principles. However, it is worth noting that there are five subjects for skin type I, nine subjects for skin type II, three subjects for skin type III, three subjects for skin type IV, and zero subject for skin type V and VI. Due to the limited sample size and

lack of data for the darkest Fitzpatrick skin types, the results attained here may not provide meaningful insight with a true accuracy of the proposed chest-based pulse oximetry on persons of all melanin levels. Note that in [59], [60], it was reported that melanin content leads to SpO_2 overestimation at low SaO_2. Further investigation is required to study the way melanin content affects the accuracy of the chest-based pulse oximetry at various SaO_2 levels.

2.4 Discussions

We unified previous evaluations of central-pulse oximetry and addressed relevant concerns while showing an accuracy that is comparable to the state-of-the-art [38]. To the best of our knowledge, this is the first thorough evaluation of chest-based pulse oximetry that jointly features a sufficient sample size, a wide dynamic range of SpO_2, minimal respiratory artifacts, and rigorous cross validation to avoid data leakage. Furthermore, the study protocol, our alignment method, and key algorithmic components were described in full detail to allow for replication. This work paves the way for realizing the simultaneous monitoring of, in addition to SpO_2, the cardiac, pulmonary, and cardiopulmonary functions using a small, standalone wearable patch device continuously and remotely, unlocking opportunities in personalized health intervention outside of a clinical setting.

2.4.1 Accurate SpO2 Estimation

We achieved low mean RMSEs for all training and calibration schemes, which were well within the criteria ($RMSE \leq 3.5\%$) for reflective pulse oximetry outlined by the FDA standard [41]. Our work addresses the challenges of the aforementioned approaches and estimated SpO_2 accurately using a novel algorithm that proves to be robust, for PPG

measured at this poorly perfused site. The breath-hold protocol successfully induced hypoxemia and reduced respiratory artifacts. Furthermore, the novel algorithms described in Section 2.4.2. to derive R leverage the morphological similarity between PPG_{red} and PPG_{IR}. Our method avoids peak and valley extraction for distorted PPG beat, and proves to be less susceptible to artifacts. The PPG_{green}-based outlier rejection algorithm was inspired by the robustness of PPG_{green} against motion artifacts [55]. Together, they alleviated difficulty in feature extraction for most PPG beats and excluded undesired PPG beats robustly.

2.4.2 Standardization of Subject-Specific Calibration

Besides accurate SpO_2 measurements, we also designed experiments to identify the best training and calibration for this dataset for improving RMSE. More data points help to better calibrate the model to the test subject, and they do so by reducing the noise in the R extracted rather than capturing a wider SpO_2 dynamic range, as evident from the results in Figure 6. Subject-specific calibration of *b* helps to reduce the randomness in the data. In contrast, if we were to calibrate *m* alone, we expect SpO_2 range to have a more important role. Ultimately, we can benefit from both longer calibration duration and wider SpO_2 range, but in a situation where hypoxemia is not preferred by the intended users, we have shown that limited SpO_2 dynamic range with a sufficiently long calibration duration can still efficiently improve accuracy. This finding is beneficial for the usage of the device since it alleviates the need to induce changes in SpO_2 and consequently makes the subject-specific calibrating procedure more practical and safer. Consequently, if necessary, breath-hold *duration* should be standardized instead of the *SpO_2 range*.

When observing the data used to calibrate both *m* and *b* for test subjects, we noticed that the first breath-hold is consistently shorter across subjects. As a direct consequence, data of the first breath-hold generally does not have a wide SpO_2 dynamic range. Furthermore, subjects seemed to be able to hold their breath longer due their adaptive tolerance to withstand the vaguely defined "discomfort" [61]. Since estimating the slope *m* requires a sufficient dynamic range of SpO_2, calibrating *m* using just the first breath-hold is usually not enough. Using all 10 breath-holds and adequate SpO_2 dynamic range across all training subjects offers a clear advantage to the globally trained *m* over the calibrated *m*. Hence, when the calibration data were limited (within the duration of one breath-hold), training *m* globally and calibrating *b* using the test subject's data can achieve better accuracy.

2.4.3 Practical Use Case

The current manufacturing cost of the device is on the order of $200. However, producing this device on a large scale can reduce the cost substantially as components would be ordered in volume and manufacturing processes can be refined to improve manufacturability. We do not foresee any challenges with scalability as the devices manufactured so far show robust functionality. The practical subject-specific calibration procedure can be designed by aggregating the conclusions made thus far. We suggest a 15s breath-hold during which the ground truth SpO_2 and biosensor data are collected from a target subject. This breath-hold duration is selected because all breath-hold durations across the subjects in this study exceed 15s. Using the data from the subjects analyzed in this study, we found the globally trained slope m_{global} to be –21.54 and the intercept b_{global} to be 106.69. Following the semi-globalized scheme, b_{global} can be replaced by $b_{\text{subject-specific}}$,

which was calibrated using data from the 15s breath-hold of the target subject. The resulting subject-specific linear model takes the form of $SpO_2 = m_{global} \times R + b_{subject\text{-}specific}$. One potential reason for calibration failure could be the adoption of smoking behavior. According to [62], smokers have elevated levels of carboxyhemoglobin (COHb). As a result, assumptions (i.e., the only hemoglobin species in the arteries are Hb and HbO_2) made in Equation (2) are violated, which can subsequently lead to the overestimation of SpO_2.

2.4.4 Limitations

One key limitation is that we only validated the accuracy of our pulse oximetry during segments with minimal respiratory artifacts. Kramer *et al.* [38] previously reported high accuracy from subjects undergoing spontaneous breathing despite a lack of details of their algorithm. Future work should investigate whether our proposed $ACr_{ed/IR}$ extraction and PPG_{green} outlier rejection algorithm can withstand respiratory artifacts more severe than the involuntary respiratory movements during breath-hold and attain similar accuracy. In addition, we noticed that the deoxygenation dynamics of R still may not be perfectly aligned to those in SpO_2, even after precise alignment. Our alignment method assumed that the delay between the start of breath-hold and chest SaO_2 deoxygenation is distributed across the integer values in the interval [–10, 10], as described in Appendix A. However, the delay may exceed beyond this, and therefore it is likely that we still captured an undesired delay for some subjects. Specifically, inter-subject and intra-subject variability in this delay may directly translate to variability in the estimated *b*, which explains the importance of calibrating *b*. For example, consider the case where the error in alignment is 1s in a breath-hold of a subject. The desaturation rate, of around 0.26% per second, can be

roughly estimated from Figure 3. The 1s error in alignment can lead to a difference of 0.26% for all datapoints of that breath-hold, and therefore, systematically introduces a bias. Besides this bias, the breath-hold study method may also lead to another shortcoming. The SpO_2 at different measurement sites may not map to one another completely, even if the delay has been accounted for [31]. This is understandable because the sternum and finger have different tissue and vascularization compositions and therefore deliver and consume oxygen at different rates as well. Finally, similar to most commercial finger pulse oximeters, the chest-based pulse oximetry may suffer from motion artifacts, the presence of Hb_{CO} and Hb_{MET}, venous pulsation [63], etc. However, we expect to considerably improve RMSE further if we can induce deoxygenation slowly and compare the estimation when using the validation protocol suggested by the FDA [41] and show improved accuracy of biosensor's measurements despite the above limitation.

2.4.5 Future Work

The results and techniques demonstrated in this work allow for the accurate measurements of SpO_2, which can in turn be used to better inform underlying pulmonary dysfunctions unobtrusively, continuously, and remotely. Together, with its ability to measure cardiac function, we can next validate the wearable patch biosensor for its ability to quantitatively and objectively assess disease progression of cardiovascular and pulmonary diseases such as COVID-19, nocturnal hypoxia caused by sleep apnea, and high-altitude sickness. Ultimately, tracking these health parameters may provide a better understanding of the cardiopulmonary-related comorbidities and consequently facilitate the adoption of longitudinal wearable monitoring devices, for detecting underlying disease when symptoms are subtle and unnoticeable.

2.5 Conclusion

In this chapter, we demonstrated that our custom, chest-worn wearable patch biosensor was capable of accurately estimating SpO_2 while subjects underwent a 10 breath-hold protocol. We presented that standardizing the calibration duration, rather than calibration range, was the most important factor for optimal calibration. Finally, we found that differences in Fitzpatrick skin types do not introduce disparities in bias. Future studies will focus on improving the study protocol to induce gradual changes in SaO_2 as per the FDA guidelines, while measuring gold standard SaO_2 simultaneously through a co-oximetry of arterial blood samples [41], designing algorithms that mitigate respiratory artifacts when present, and by recruiting a larger population that is demographically diverse, especially participants with higher Fitzpatrick skin types. Together with its holistic cardiac monitoring, this device can provide longitudinal and quantitative information of disease progression in both cardiovascular and pulmonary diseases.

By replacing a functional module of a classical DSP pipeline in Section 2.2.4.3, we can improve the accuracy of the target task by using the multiple datapoints available instead of just two of them. Because most of the DSP pipeline still followed the principle of pulse oximetry, this approach has improved the accuracy without much risk—we still can peek into the modules prior and after our proposed data-driven feature extractor for debugging and data quality assessment. We will carry out the same strategy in **CHAPTER 3**, in which our objective was to estimate RR from ECG and SCG.

CHAPTER 3. USING U-NET DENOISING FOR RESPIRATORY RATE ESTIMATION DURING EXERCISE RECOVERY

At-home monitoring of respiration is of critical urgency especially in the era of the global pandemic. ECG and SCG—measured in less cumbersome contact form factors than the conventional sealed mask that measures respiratory air flow—are promising solutions for respiratory monitoring. RR can be estimated from EDR and SDR signals. However, non-respiratory artifacts might still be present in these surrogates of respiratory signals, hindering the accuracy of the RRs estimated.

Similar to the strategy in **CHAPTER 2**, we approached this problem by first adopting the classical DSP pipeline to estimate RR from ECG and SCG, followed by replacing a functional module with a data-driven approach. Specifically, we used a U-Net DL architecture to denoise the EDR and the SDR signals extracted. U-Net was originally developed for segmenting biomedical images, but it was later used in other tasks: image sharpening for medical images, source separation for decomposing music audio signals into vocal and backing track components, etc. The diverse range of applications of U-Net sparked our interest in using it in this work. Upon inspecting the Short-Time Fourier Transform (STFT) of these derived respiratory signals as 2D images, the selection of U-Net thus became a natural choice. Denoising EDR and SDR signals can be considered as sharpening task or the separation of respiratory and non-respiratory components in their STFT 2D images. We also noticed that the patterns in the STFT 2D images are translational invariant, so the trained U-Net may generalize well to the different frequency range of respiration. In other words, by applying U-Net on the STFT of the EDR and the SDR signals, we can extrapolate its denoising capability to RR.

3.1 Introduction

The current clinical workflow lacks the ability to detect and monitor pulmonary diseases at home. As a result, due to the recent novel COVID-19, clinicians were unequipped to capture the subtle increase in resting RR of patients beginning days prior to diagnosis [9], [64]. One effective way to prevent disease spread is to monitor RR remotely for subjects who are under high infection risk such as those in close proximity to patients diagnosed with COVID-19. In addition, presymptomatic detection of COVID-19 infection can assist intervention strategies in a timely manner.

3.1.1 Respiratory Rate Estimated from Wearable Patch

The gold standard for RR estimation is through the measurement of respiratory air flow, but this approach requires a sealed mask or cannulas to be worn and is thus obtrusive for normal daily living [20]. Accordingly, researchers are investigating solutions for estimating RR without the need for such cumbersome equipment to be placed over the nose or mouth of the user. Respiratory effort belts are unobtrusive, yet there are other alternatives that are less limited by discomfort to the patient and low signal quality. Non-contact sensors are capable of estimating RR accurately [65], but they may have lower signal quality compared to contact sensors. Therefore, contact sensors are more suitable for at-home monitoring of RR. In particular, wearable technologies at the chest, wrist, or finger—that capture physiological signals—hold promise for unobtrusive remote respiratory monitoring. Specifically, respiratory information are derived from the non-respiratory measurements taken by such wearables, including the ECG [66]–[69], PPG [70]–[72], SCG , or combinations of these measures [75]–[79].

Our custom wearable patch has previously shown promise in measuring cardiovascular function in multiple applications, and thus it provides accurate and robust measurement of ECG and SCG signals [33], [35], [80], [81]. This combination of the measurement site and signals captured is not only beneficial to pulmonary disease

monitoring, but also crucial toward understanding cardiopulmonary interactions [82]. Moreover, the placement of the patch on the sternum as compared to the wrist, arm, or other peripheral sites on the body can reduce the impact of motion artifacts on the recordings, since the torso moves less in normal daily activities than the extremities.

An important innovation in our approach is the combination of ECG and SCG signals for deriving RRs: we anticipate that the combination of ECG and SCG modalities will yield a more accurate and robust estimation of RR when compared to single modality. The ECG captures the electrical activity of the heart while the SCG captures the mechanical activity. These signals are thus modulated by respiration through different mechanisms and to different extents. Consequently, they are affected by motion in different manners, potentially allowing their combination to be synergistic toward yielding a high-quality RR estimate. Respiratory activities are translated to signals through amplitude modulation (AM), frequency modulation (FM) and baseline wander (BW) [20]. In this work, we primarily focus on AM-based methods and do not use FM-based or BW-based demodulating techniques due to their poorer representative power of respiration.

3.1.2 Denoising Demodulated Respiratory Signals

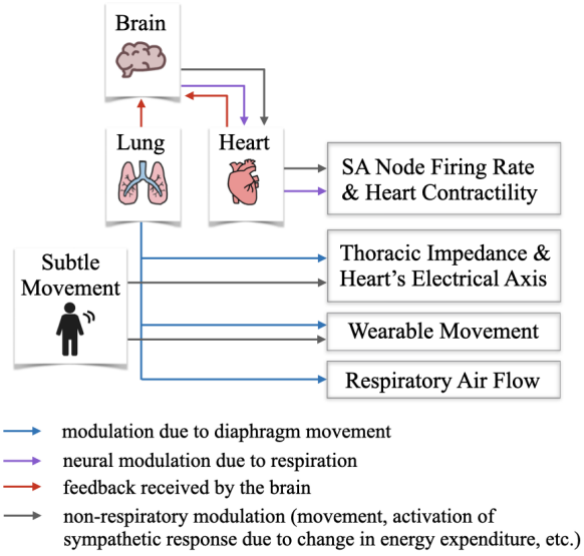


Figure 3.1. Respiration can modulate physiological signals collected from a chest-worn wearable patch differently. Simultaneously, non-respiratory artifacts can interfere with this phenomenon, increasing difficulty to derive a clean respiratory signal that resembles respiratory air flow.

AM scheme modulates the amplitude of the QRS complex—associated with ventricular depolarization—of the ECG [83] and the morphology of the systolic profile of the SCG [17]. More details of the AM-based demodulation techniques used in this work are provided in Section 3.2.3. These demodulating techniques, however, are limited to controlled environments where non-respiratory modulation is minimal. As illustrated inFigure 3.1, there are other sources that similarly modulate the measured physiological signals. For example, subjects may have subtle movements due to involuntary muscle tremors even when they are requested to remain still. These motion artifacts cannot be easily removed by digital filtering when the frequency content of the movement overlaps with that of ECG and SCG [84]. Finally, aliasing due to nonuniform sampling of AM (non-stationarity of HR) can further challenge RR estimation [85]. These non-respiratory artifacts can be non-linear and cannot be mitigated with demodulating techniques alone.

When the corruption to the physiological signals persist in the demodulated respiratory (DR) signals, non-respiratory artifacts can still be reduced by DSP methods

such as modeling of temporal dependencies, fusion techniques, and outlier rejection based on quality assessment [20]. These methods are usually applied as separate blocks, which require a cumbersome hand-designed (manual) process of each individual block and could sometimes discard key respiratory information. Although this process provides insight on how well each block performs, such increased complexity may obfuscate respiratory information.

DL is one solution to address the limitations of the DSP methods. One common and relevant application of DL is to learn a nonlinear mapping to remove artifacts. The U-Net, a DL architecture originally designed for biomedical image segmentation [86], combines feature maps extracted from the contractive and the expansive paths through successive convolutional layers through the skip connections. The U-Net has later been adapted for an audio source separation problem for its capability to reconstruct spectrograms that preserve high-level detail in both spectral and temporal domains [87]. In a preliminary analysis, we noticed that the non-respiratory artifacts could corrupt spectrograms of the DR signals in ways comparable to those in audio source separation problems. By formulating denoising DR signals as a source separation problem, i.e., respiratory activity as one source and non-respiratory activity as the other, we can leverage the U-Net and address non-linearity mapping, temporal dependencies modeling, and fusion in a data-driven manner. A data-driven approach has the adavantage of adapting to different applications if the training and the testing data share similar distributions.

In this chapter, we innovatively estimated RR from a combination of single-lead ECG and triaxial SCG recorded by our custom wearable patch placed mid-sternum. To reduce non-respiratory artifacts, we proposed a cascaded framework that includes a demodulating block for exploiting respiratory modulation and a denoising block for reducing non-respiratory artifacts. Finally, we examined the representative power of the

DR inputs, the denoising capability of the U-Net, improvements by fusion, and the generalizability of the proposed cascaded framework.

3.2 Methods

3.2.1 Dataset

The dataset used here was collected in a previous study described in detail by Shandhi *et al.* in [35]. The protocol (H18452) was approved by the Georgia Institute of Technology Institutional Review Board. The dataset has been de-identified and is used here for post-hoc analysis only.

In this dataset, we recruited 21 healthy subjects and analyzed 17 (Age: 26.8±4.1 years, Weight: 67.5±14.1 kg and Height: 170.5±9.9 cm) of those whose data have better physiological signal quality as determined by visual inspection. The poor signal quality, defined as ECG segments without distinctive QRS characteristics, was likely due to the detachment of the ECG electrodes during exercise. This dataset includes high fidelity single-lead ECG and tri-axial SCG signals recorded using a wearable patch that was placed on subject's mid-sternum. Since the mechanical activity of the heart is mostly captured by the dorsal-ventral (D-V) direction [81], we only used $SCG_{D\text{-}V}$ in this work. For simplicity, we will refer to $SCG_{D\text{-}V}$ as SCG hereinafter. In parallel, gold standard RR and other metabolic labels were collected using a COSMED K5 wearable metabolic system (COSMED, Rome, Italy), sampled at each breath. This K5 system reported limits of agreement of 0.30±2.67 bpm for their RR measurements [88]. The error may be due to the metabolic artifacts in the K5 system. The subjects were asked to perform several tasks such as treadmill walking, outdoor walking, and resting. In this study, only data associated with the 5-minute resting period after treadmill walking (treadmill recovery) were used for

training and testing. The data from the wearable patch and the K5 system were synchronized prior to the subsequent analysis.

During treadmill recovery, subjects stood still and breathed naturally. The RRs collected during this period range from 5-40 bpm (equivalently 0.083-0.67 Hz). The wide dynamic range of RR measurements in this dataset renders it feasible towards generating an algorithm to estimate RR in subjects who experience pulmonary dysfunction, which usually results in elevated resting RR.

3.2.2 Cascaded Framework

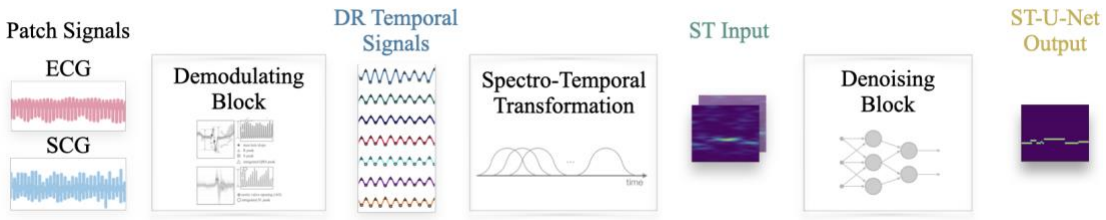


Figure 3.2. An overview of the proposed cascaded framework. Seven respiratory signals were derived from ECG and SCG according to amplitude modulation (AM). The time domain signals were transformed to spectro-temporal (ST) representations to reduce data dimensionality and to retain information that are most relevant to the context. Finally, a U-Net was employed to achieve temporal dependencies modeling, nonlinear mapping, and sensor fusion to reduce non-respiratory artifacts.

The overview of the cascaded framework is illustrated in Figure 3.2. The proposed cascaded framework combines the advantages of both demodulation and U-Net, which will be referred to as the demodulating block and the denoising block, respectively. Several demodulating techniques were employed to extract DR signals. The resulting DR signals were transformed to spectrograms and segmented with overlapping windows to produce ST representations. The labels (also transformed to ST representations) and the outputs of U-Net were used to compute the loss for learning the parameters of the U-Net. The dominant frequency was then selected as the frequency with highest energy at the temporal

step corresponding to the label. The dominant frequency can be translated to RR through unit conversion.

3.2.3 Demodulating Block

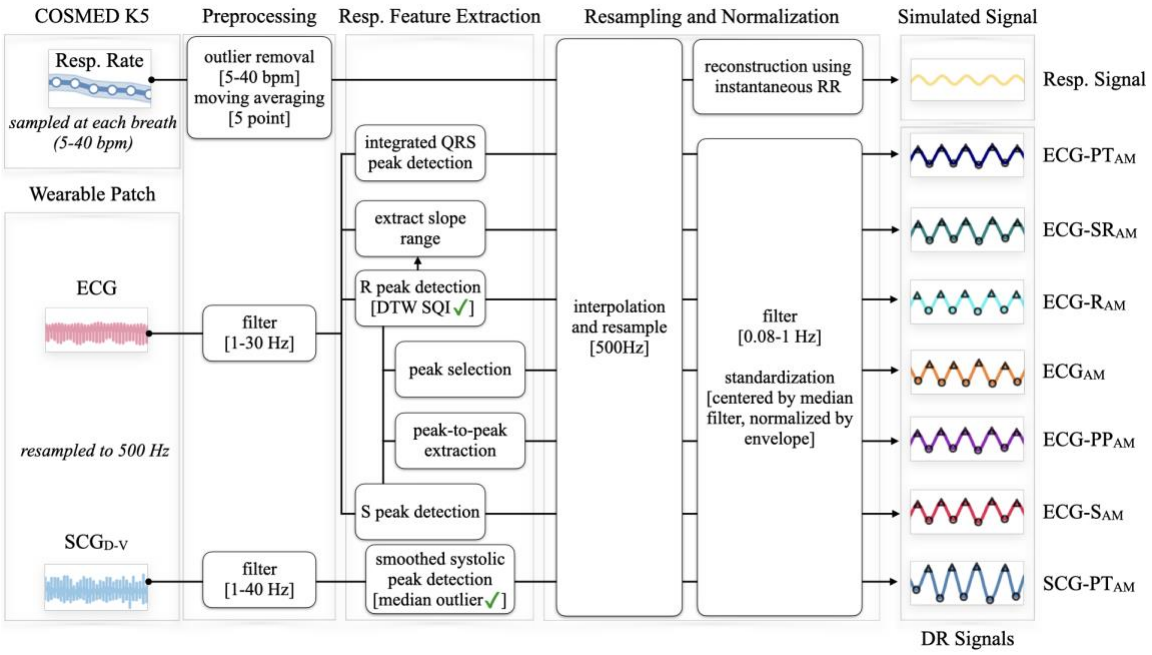


Figure 3.3. Block diagram of the demodulating block. The physiological data, collected from the wearable patch and COSMED K5 system, were transformed to pulsatile demodulated respiratory (DR) signals. Physiological signals were first preprocessed with filters. Next, several features representing the respiratory activity were extracted. These features were subsequently resampled and normalized to produce six ECG-derived respiratory (EDR) signals and one SCG-derived respiratory (SDR) signal. The peaks (triangles) and valleys (circles) of the DR signals extracted were used to compute a benchmark for RR estimation (i.e., a temporal method).

In this section, successive steps to demodulate respiratory information were described. As illustrated in Figure 3.3, we first preprocessed the physiological signals to attenuate artifacts. Next, features that were representative of respiratory information were extracted. These features were interpolated and filtered to construct evenly sampled, smooth DR signals.

3.2.3.1 Preprocessing

Outliers of RR labels exceeding or below the normal range of 5-40 bpm, potentially due to measurement error by the COSMED K5 system, were first removed. The RR labels were smoothed with a 5-point moving average (MA) filter to reduce noise. For signals from the wearable patch, ECG was filtered by a 4^{th} order Butterworth filter with a passband of 1-30 Hz. SCG was filtered by a 4^{th} order Butterworth filter with a passband of 1-40 Hz. The cutoff frequencies were selected to preserve relevant components of the signals (e.g., QRS complex and systolic profile) and to remove out-of-band noises.

3.2.3.2 Respiratory Feature Extraction

Once signals were filtered, we rejected ECG beats with low SQIs. An ECG template was first constructed by averaging the ECG beats for each subject. An SQI algorithm based on dynamic time warping (DTW) distance described in [89] was adopted, which has a range from 0 to 1. Note this DTW-based SQI algorithm was first introduced and validated for PPG signals [90]. We rejected ECG beats and the corresponding SCG beats with SQIs lower than a threshold of 0.5. This SQI algorithm was qualitatively assessed and validated using this dataset. On average, only 0.06% of the ECG beats per subject were rejected.

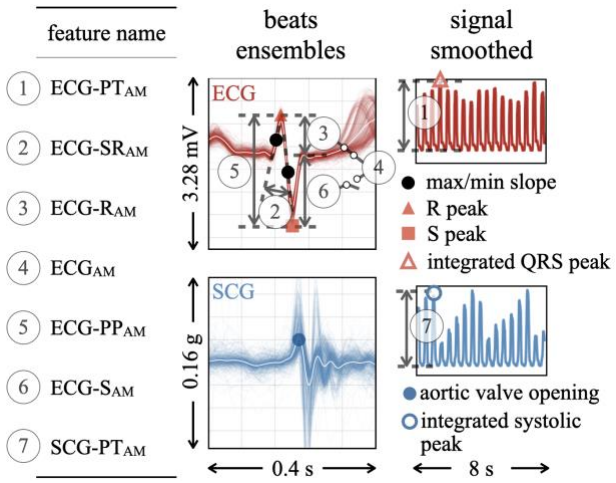


Figure 3.4. Different features of ECG and SCG extracted. In this work, only ECG and SCG signals demodulated using AM were presented. ECG-PT_{AM} is the amplitude of the smoothed QRS complex. ECG-SR_{AM} is the slope range amplitude. ECG-R_{AM} is the R peak amplitude. ECG_{AM} is either the R peak or the S peak (subject-dependent) amplitude. ECG-PP_{AM} is the QRS peak-to-peak amplitude. ECG-S_{AM} is the S peak amplitude. SCG-PT_{AM} is amplitude of the smoothed SCG systolic profile.

We used the remaining ECG and SCG beats to extract seven AM features as an intermediate step of the demodulating block (as shown in Figure 3.4). QRS complexes were detected using the Pan-Tompkins algorithm [49], followed by identification of R peaks (ECG-R_{AM}) and S peaks (ECG-S_{AM}). QRS peak-to-peak (ECG-PP_{AM}) was calculated from the absolute difference between an R peak and an S peak. The timestamp of each ECG-PP_{AM} feature is the average timestamps of R and S peaks. Slope range (ECG-SR_{AM}) was also extracted as described in [69] for its robustness against noise. Smoothed QRS peaks (ECG-PT_{AM}) were extracted using the intermediate output generated from the Pan-Tompkins algorithm. We also designed a subject-dependent switch to select between R peak and S peak. This switch was implemented to determine the peak that has higher absolute amplitudes on average (ECG_{AM}). As inspired by the Pan-Tompkins algorithm [49], SCG was squared, integrated, and smoothed. The peak values at the smoothed systolic profile of each cardiac cycle were extracted (SCG-PT_{AM}). To remove outliers in SCG-PT_{AM}, we first derived a reference array (same size as the SCG-PT_{AM}) using a median filter

for each subject. SCG-PT$_{AM}$ features deviate too much (five times the median of the features) from the reference array were rejected. On average, only 1.07% of the SCG-PT$_{AM}$ features were rejected per subject. Note that no ECG features were rejected.

We chose different outlier rejection algorithms because the one for ECG beats is more suitable for comparing signals and the other one for SCG-PT$_{AM}$ is more suitable for detecting outliers in a distribution of features. Since each of these respiratory features can have a different respiratory SNR under different situations, they may provide complementary information to each other.

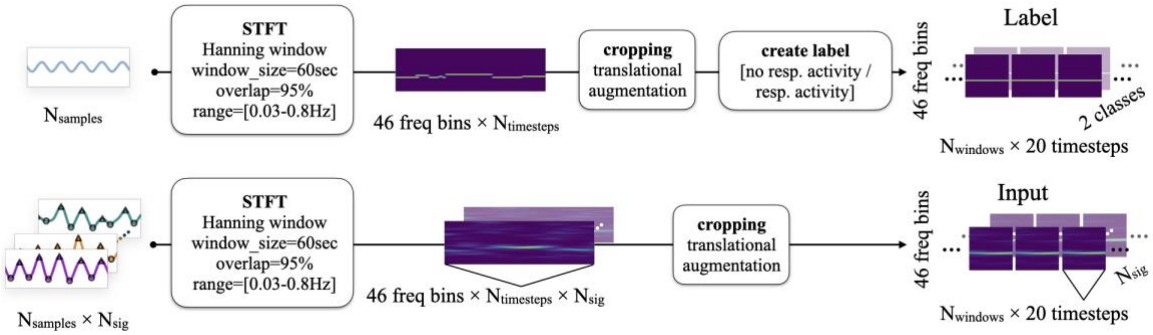


Figure 3.5. Transformation of the DR signals to the ST representations. Short-Time Fourier Transform (STFT) was applied to first produce long ST representations. These ST representations were cropped to achieve translational augmentation, resulting in data with a dimension of 46 × 20 × N_{window} for each DR signal and each subject.

3.2.3.3 Resampling and Normalization

Prior to transforming the DR signals to ST domain, we removed a portion of the non-respiratory artifacts using filters and smoothing methods. The aforementioned AM features were sampled at each cardiac cycle. Since HR is time varying, it is necessary to resample the features before filtering. The data were interpolated to 500 Hz using linear spline then smoothed using a MA filter with a 1-second window. We selected this approach over the cubic spline interpolation to avoid the artifacts near the edges. The MA filter was

selected for its simplicity in implementation and its ability to remove interpolation-induced artifacts particularly in the SCG-PT_{AM} signal. The 1-second window was empirically determined and verified by visual inspection of the DR signals. Specifically, we used the basic principles of MA filter to fine-tune this parameter: 1) if high frequency noises that are unrelated to respiration are observed after applying this filter, the window size might be too small, and 2) if respiratory signals are not visible after applying this filter, the window size might be too large. Though demodulated signals were usually resampled to 4 Hz, we chose 500 Hz for the purpose of seamless inspection in conjunction with ECG and SCG. This may add computational complexity—unsuitable for real-time estimation—but we observed that transitioning from 500 Hz to 4 Hz is straightforward and does not affect the conclusions made in this work. Next, the smoothed signals were filtered by a 4^{th} order Butterworth filter with a passband of 0.08-1 Hz. The resulting DR signals were pulsatile waves containing mostly respiratory information. The peaks and valleys of the DR signals (on the right of Figure 3.4) were extracted. Despite the success of applying the demodulating techniques, these signals can still inherently suffer from physiological dynamics. Specifically, the baseline drift may still be present (those overlapped the frequency range of respiration cannot be removed by filtering), and the peak-to-peak amplitudes of these pulsatile waves can still vary over time (due to changes in tidal volume). To mitigate these problems, we heuristically applied a median filter of window size 2001 (equivalent to 4 seconds) extract the baseline drift for centering. The centered DR signals were then rescaled to minimize the effect of time-varying tidal volume. Rescaling is performed by dividing the centered signal with signal envelope extracted from

Hilbert transform, so most of the signals exhibited a range of [-1, 1]. The operation to normalize a DR signal for a subject can be summarized as follow:

$$x_{i,normalized} = \frac{x_i - x_{i,drift}}{x_{i,HilbertEnvelope}} \tag{3.1}$$

where x_i is the resampled DR signal, $x_{i,drift}$ is its baseline drift, $x_{i,HilbertEnvelope}$ is the envelope of $x_{i,drift}$, and $x_{i,\ normalized}$ is the normalized DR signal. Segments of the signal exceeding [-1.5, 1.5] were clipped. Due to the non-linearity of the median filter, it could potentially introduce artifacts that are unrelated to respiration. Fortunately, such events were not common, and we expect the denoising block to mitigate these artifacts. These rescaled DR signals are unitless.

3.2.3.4 Overview of CardioRespNet and Its Functional Blocks

Since the RRs provided by COSMED K5 were coarsely sampled at every breath, we lacked the continuous, pulsatile respiratory signal that resembles respiratory air flow. We related the two by treating RRs as a message signal that modulated the phase of the respiratory signal. RR samples were first interpolated to 500 Hz. The resampled RR, or equivalently the instantaneous frequency $f_{instantaneous}(t)$, was then converted to the phase $\phi(t)$ through integration, as shown in Equation (**3.2**). This integration was approximated by the Trapezoid rule in implementation. Finally, $\phi(t)$ was next used to modulate the phase of the carrier signal with an arbitrary frequency f_c per Equation (**3.3**). The resulting reconstructed signal $y(t)$ is a pulsatile wave with a range of [-1, 1].

$$\phi(t) = 2\pi \int_0^t f_{instantaneous}(\tau)\, d\tau \tag{3.2}$$

$$y(t) = sin(f_c + \phi(t)) \tag{3.3}$$

3.2.4 Spectro-Temporal Transformation

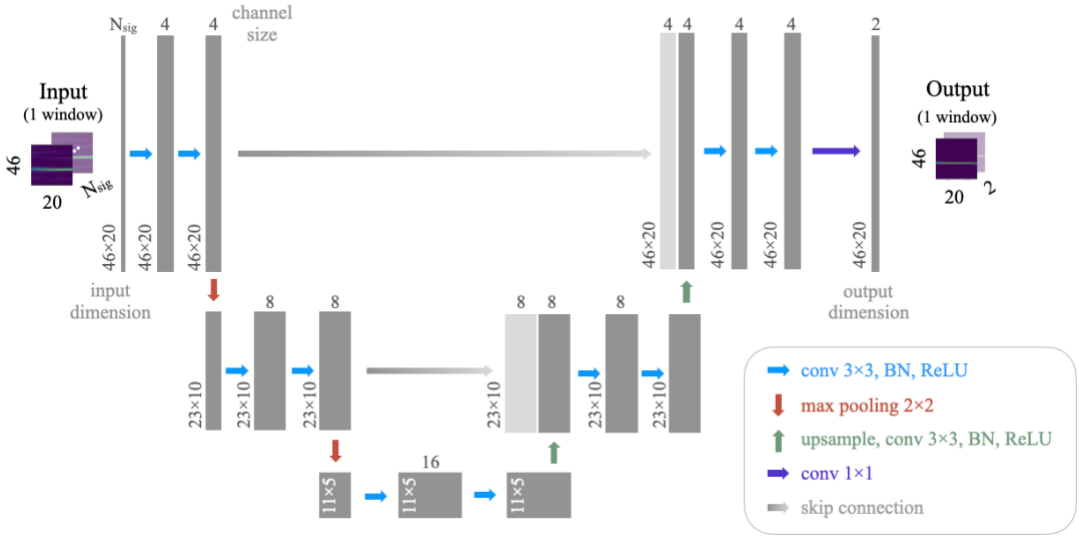


Figure 3.6. The denoising block: U-Net architecture designed to reduce non-respiratory artifacts. Data dimension after each operation is denoted at the bottom left of the resulting feature map (gray rectangle). The number of channels is denoted on the top of the feature map. In the most compact form of the data, it has a dimension of 11 × 5 pixels (11 temporal steps and 5 frequency bins). Each colored arrow indicates a different operation as denoted in the text box on the right. BN is the acronym for batch normalization. ReLU is the acronym for Rectified Linear Unit.

Next, we transformed the demodulated, normalized signals to spectro-temporal (ST) domain by STFT. The window size is 60 seconds with 95% overlap. Hanning window was selected as the window function to soften artifacts. The resulting spectrogram for a DR signal has a dimension of 250 × $N_{timestep}$, where 250 corresponds to the number of frequency bins (half of the sampling rate) and $N_{timestep}$ is the number of timesteps (varied by the length

of treadmill recovery recorded). The spectrograms were truncated in the spectral dimension to preserve content (0.03-0.8 Hz) just slightly wider than that of respiration. As a result, the data were reduced to a dimension of $46 \times N_{timestep}$. Finally, the truncated spectrograms for each subject were cropped to introduce translational augmentation. This was performed by applying a moving window of 20 timesteps with 19 overlapping time steps. In other words, ST data with a dimension of $46 \times N_{timestep}$ were further cropped to produce ST data with a dimension of $46 \times 20 \times N_{window}$ for each subject and each DR signal, where N_{window} is the number of moving windows applied. Each window has a spectral resolution of 0.0167 Hz (equivalent to 1 bpm) and a temporal resolution of 3 seconds.

3.2.5 U-Net Architecture

This implementation attempt of the U-Net was similar to that described in the original U-Net paper by Ronneberger *et al.* [86], but with a slightly different hyperparameter set. The U-Net presented in this work consists of a contractive path, a bottleneck section, and an expansive path. The contractive path focuses on learning a compact representation of the input data, which have a dimension of $N_{mini\text{-}batch} \times N_{sig} \times 46 \times 20$, where N_{sig} indicates the number of DR signals used and $N_{mini\text{-}batch}$ is the number of windows used during mini-batch gradient descent. Note that the order of the dimensions can be directly mapped to the convention used in the 2D convolution implemented by PyTorch (number of images × number of input channels × height of an image × width of an image).

As illustrated in Figure 3.6, there are two repeated contractive operations. Each operation consists of two convolutional blocks (blue arrows), followed by a max pooling

layer (red arrows). Each operation reduced each data dimension by half but doubled the channel size. The bottleneck section received the outputs of the contractive operations and expanded the channel size to 16. In the expansive path, operations mirroring the ones in the contractive path were performed. Max pooling layers were replaced by up-convolutional layers (green arrows), followed by two convolutional blocks (blue arrows). Each expansive operation doubled each data dimension and halved the channel size. The contractive path and the expansive path communicated not only through the bottleneck layer but also through skip connections (gray arrows). Due to the loss of the border by up-convolution, the feature map in the expansive path has a slightly smaller dimension than its corresponding counterpart in the contractive path. As a result, padding was performed on the expansive feature maps. Finally, a 1 × 1 convolutional layer (purple arrow) was applied to reduce the channel size to 2. These two dimensions represent the two classes each pixel belongs to (0 for non-respiratory pixels and 1 for respiratory pixels). The convolutional and up-convolutional blocks used 3 × 3 kernel, strides of 1, and padding of 1. The max-pooling layer used 2 × 2 kernel, strides of 2, and padding of 0. Batch Normalization (BN) layer was applied to accelerate training [91] and a Rectified Linear Unit layer was used as the activation function.

3.2.6 Fusion

Fusion by U-Net was achieved at the first convolutional layer, where the filtered outputs of all DR inputs were summed to produce a single feature map for each convolutional filter. Additional DR inputs can be easily incorporated into the U-Net by adjusting the number of input channel. For comparison, a simple fusion method was also

implemented by averaging RRs estimated from the DR inputs, which is found to be more accurate than the spectral averaging method described in [20] for this dataset.

3.2.7 Loss Function

To learn the parameters of U-Net, we employed both the binary cross entropy (BCE) loss, $loss_{BCE}$, and a soft Dice loss, $loss_{Dice}$ [92], [93]. They can be expressed as follow:

$$loss_{BCE} = -\frac{1}{N}\sum_{i=1}^{N} y_i log(p_i) + (1-y_i) log(1-p_i) \tag{3.4}$$

$$loss_{Dice} = 1 - \frac{\sum_{i=1}^{N} y_i p_i}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} p_i + \varepsilon} \tag{3.5}$$

where N is the number of samples in each mini-batch, p_i is the prediction, and y_i is the corresponding reference label (p_i and y_i are both in the ST domain). BCE loss was selected since it is the standard loss function for most binary classification problems. Dice loss was selected to mitigate class imbalance problem in the dataset. Due to the compressive nature of the STFT, the respiratory contents were sparse relative to the non-respiratory ones. Hence, the classes of the pixels are very imbalanced. To ensure the loss is defined in all cases, a smoothing factor was added to the denominator of Dice loss. These two losses contributed equally to the overall loss:

$$loss = 0.5 \times loss_{BCE} + 0.5 \times loss_{Dice} \tag{3.6}$$

3.2.8 Training and Testing Model

To evaluate the performance of the U-Net, the subjects were randomly split, by a ratio approximately equal to 6.5:3.5, into a training (N=11) and testing (N=6) dataset based on their subject ID. Five-fold cross validation (again, split by subject ID) repeated 5 times was performed on the training dataset to search for the optimal hyperparameter set, defined as the one with the lowest average mean absolute error (MAE) on the validation partitions. The hyperparameters include the loss function, model depth, channel size, and the DR inputs selected for fusion. An optimal model defined by the optimal hyperparameter set was trained by the training dataset, and only results on the testing dataset were reported. During backpropagation, parameters were updated using Adam optimizer with a learning rate of 0.001 and a weight decay of 0.01. The mini-batch size was set to 8. All models were trained for 50 epochs.

When sampling mini-batches from the training dataset, data augmentation techniques including superpixeling, Gaussian blurring, addition of salt/pepper noise, and identity transformation were employed to introduce artifacts that resembled non-respiratory activities expected in the ST domain. These data augmentation techniques were all applied independently for each DR input. In each mini-batch, one of these augmentation techniques was selected with equal probability to transform the data.

3.2.9 Performance Evaluation

To assess the proposed framework, we computed the MAEs on the testing dataset using the temporal method, the ST method, and the ST-U-Net (combining ST transformation with the denoising U-Net) method. They are the outputs of the demodulating block, ST transformation, and the denoising block respectively (see Figure

3.2). The temporal method estimates RR by dividing the total number of peaks and valleys of the DR signals, within 30 seconds preceding and succeeding (60 seconds in total) each RR label, by two. Both the ST method and the ST-U-Net method select the dominant frequency as the estimated RR. Although the loss function helped evaluate how well the model learned, denoising capability of the U-Net could be more intuitively assessed by MAEs. We also computed the R^2 of the best model. A few qualitative analyses were performed to give insights on the improvements achieved by the U-Net. This is accomplished by showing ST representations scaled by min-max normalization to the range of [0, 1].

3.3 Results

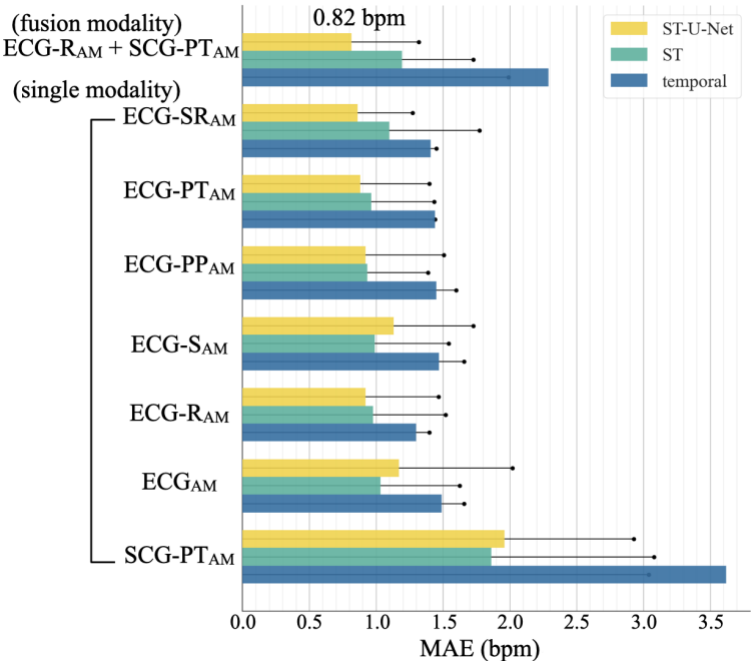


Figure 3.7. Mean absolute errors (MAEs) of the temporal, ST, and ST-U-Net methods are indicated by the blue, green and yellow color bars respectively. The standard deviations of the absolute errors are indicated by the black lines. EDRs are more accurate than SDR in estimating RR. However, by fusing ECG-RAM and SCG-PTAM, the proposed method achieved the lowest MAE at 0.82 bpm among all choices of inputs (seven single modalities and one fusion modality).

Figure 3.7 reports the MAEs of eight choices of DR inputs, which include seven single modalities and one fusion modality. Both the ST method and the ST-U-Net method appear to be superior to the temporal method. In particular, when comparing the ST and the ST-U-Net methods to the temporal method, the percentage changes in mean MAEs across all DR choices were considerably lower, in the range of [-22.14, -48.53] % and [-21.30, -64.30] % respectively. Next, we compared the ST method and the ST-U-Net method. For single modalities, all six EDRs (ECG-RAM, ECG-PTAM, ECG-SAM, ECG-PPAM, ECGAM, ECG-SRAM) achieved lower MAEs than the SDR (SCG-PTAM) prior to denoising. After denoising the data with U-Net, MAEs of EDRs remained lower than that

of SDR. Among these single modalities, denoised ECG-SR_{AM} achieved the lowest MAE of 0.86 bpm, though its MAE was not the lowest prior to denoising. Generally, those with lower MAEs (ECG-R_{AM}, ECG-PT_{AM}, ECG-PP_{AM}, and ECG-SR_{AM}) could benefit from the U-Net's denoising capability. The changes in MAEs were -0.06 bpm (-5.67 %), -0.08 bpm (-8.45 %), -0.01 bpm (-1.39 %), and -0.24 bpm (-21.36 %) respectively. Yet, the U-Net could incur higher MAEs for the three DR inputs (ECG_{AM}, ECG-S_{AM}, and SCG-PT_{AM}) with the highest MAEs prior to denoising. The changes in MAEs were +0.14 bpm (+13.56 %), +0.14 bpm (+14.47 %), and +0.1 bpm (+5.24 %) respectively.

For the fusion modality, we also reported the MAEs of ECG-R_{AM} and SCG-PT_{AM} using the temporal, the ST, and the ST-U-Net methods. In the subsequent sections, we will refer to the U-Net that fused ECG-R_{AM} and SCG-PT_{AM} as the fused U-Net. Although MAE achieved by the denoised SCG-PT_{AM} alone was higher than any of the denoised EDRs, fusing ECG-R_{AM} and SCG-PT_{AM} using U-Net led to the optimal MAE of 0.82 bpm, which is lower than MAEs achieved by all single modalities. A R^2 of 0.89 was achieved. As expected, we found that the MAE of the simple fusion technique (1.20 bpm) was between that of SCG-PT_{AM} (1.86 bpm) and ECG-R_{AM} (0.98 bpm) prior to using U-Net. The change in MAE was -0.38 bpm (-31.61 %) after applying the U-Net, which showed that the U-Net also improved RR estimation in fusion modality.

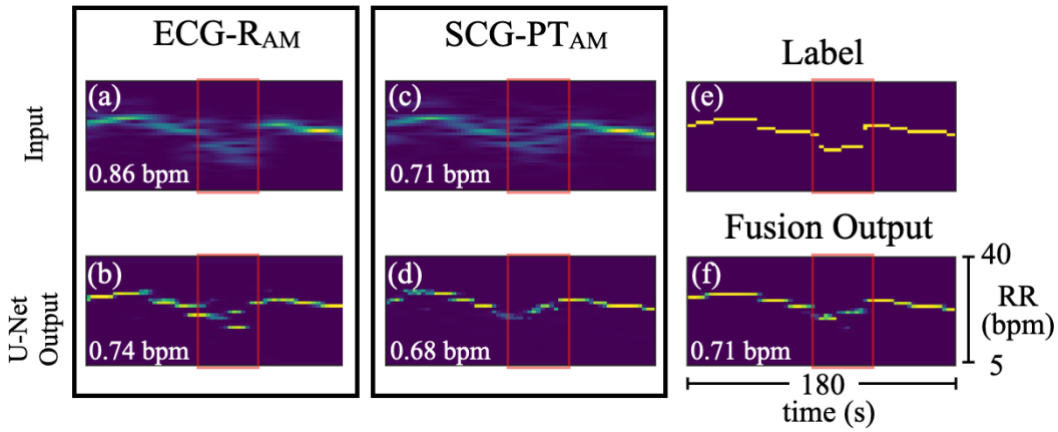


Figure 3.8. ST representations of input, U-Net output, and label for subject seven. MAEs are provided at the bottom left corners. A segment (red box) of the fused U-Net output shows that complementary information of ECG-R$_{AM}$ and SCG-PT$_{AM}$ enhanced respiratory content since it showed great agreement with the label. Although MAE of fused U-Net is higher than that of the denoised SCG-PT$_{AM}$ (single modality), quantitative analysis on all test subjects suggests the superiority of fusion over single modality. (a) ECG-R$_{AM}$ input. (b) U-Net output of ECG-R$_{AM}$. (c) SCG-PT$_{AM}$ input. (d) U-Net output of SCG-PT$_{AM}$. (e) label. (f) U-Net fusion (ECG-R$_{AM}$ and SCG-PT$_{AM}$) output.

As depicted in Figure 3.8, we qualitatively showed that the U-Nets, either of single modality or fusion modality, successfully reduced non-respiratory artifacts for subject seven. Prior to denoising, a segment (red box) of the input of ECG-R$_{AM}$ and SCG-PT$_{AM}$ appeared to be blurry. Fortunately, applying the fused U-Net reduced the non-respiratory artifacts (blurriness) and sharpened the respiratory content (dominant spectral components in the ST domain). Results from two U-Nets of single modality trained by ECG-R$_{AM}$ (Figure 3.8 (a)) and SCG-PT$_{AM}$ (Figure 3.8 (c)) separately were also presented. Their denoised output (Figure 3.8 (b) and (d)) were in close agreement with RR label (Figure 3.8 (e)), but the blurriness was still somewhat observable in the aforementioned segment. By applying a fused U-Net, we obtained denoised output (Figure 3.8 (f)) that were in closer agreement with the label (Figure 3.8 (e)). Not only did the location of the respiratory content matched that of the labels in the ST domain, but also it exhibited higher amplitude overall.

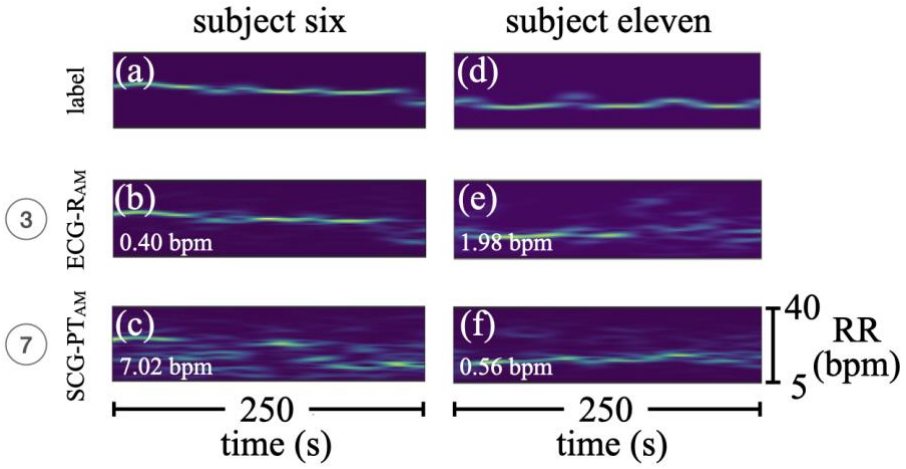


Figure 3.9. ECG and SCG can couple with respiratory activity to different extents due to subject variability. The MAEs are shown in the bottom left corners. (a) ST representation of the reference label for subject six. (b) ST representation of ECG-R_{AM} for subject six. (c) ST representation of SCG-PT_{AM} for subject six. (d) ST representation of the reference label for subject eleven. (e) ST representation of ECG-R_{AM} for subject eleven. (f) ST representation of SCG-PT_{AM} for subject eleven.

Figure 3.9 shows the two DR inputs ECG-R_{AM} and SCG-PT_{AM} in ST domain for subject six and eleven. Although in general, EDRs exhibited stronger coupling with respiration than SDR did, SDR could occasionally be less corrupted by non-respiratory artifacts. For subject six, ECG-R_{AM} was clean throughout the recording, but SCG-PT_{AM} was severely corrupted. For subject eleven, we observed that SCG-PT_{AM} has higher SNR than ECG-R_{AM} in the second half of the 250 s recording. The MAEs presented in Figure 3.9 also suggest that these signals may have varying representative power of the respiratory activity across subjects.

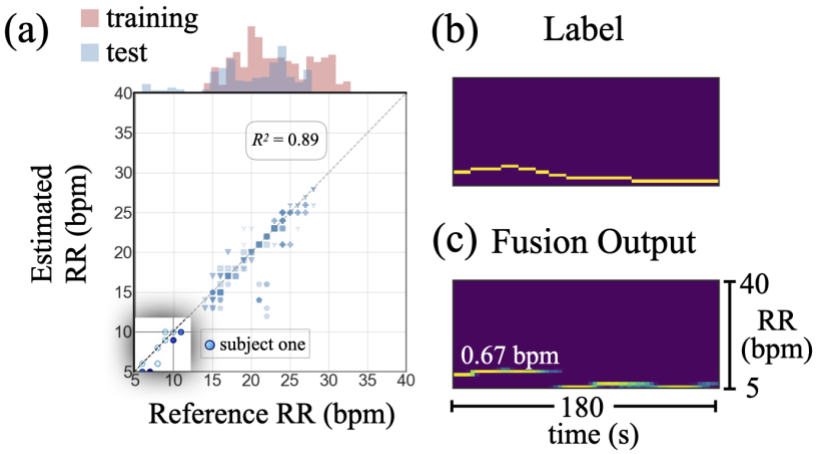


Figure 3.10. The fused U-Net can be extrapolated to unseen RRs. (a) Correlation analysis on all subjects in the testing dataset. The distributions of training (red) and testing datasets (blue) are shown on the top of the scatter plot to help visualize the discrepancies in RR distributions. In the scatter plot, only testing data are presented. A coefficient of determination (R^2) of 0.89 is reported. (b) Label in ST domain for subject one. (c) Fusion output in ST domain for subject one. The MAE for subject one was 0.67 bpm.

Figure 3.10 illustrates that the denoising capability of the fused U-Net could be extrapolated to a range of RRs that was unseen in the training dataset. As shown in Figure 3.10 (a), RRs for subjects in training dataset were all above 12 bpm, and RRs for one of the testing subjects (subject 1) were below 12 bpm. When inspecting the output of the fused U-Net for this subject—who had an MAE of 0.67 bpm after denoising—the respiratory content recovered (Figure 3.10 (b)) was in great agreement with that of the label (Figure 3.10 (c)).

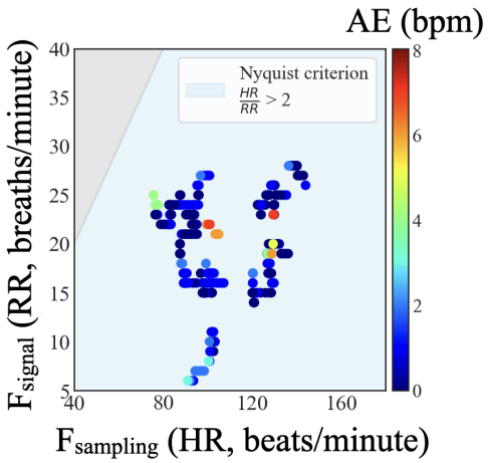


Figure 3.11. All data in the testing dataset satisfied the Nyquist criterion. Most data close to Nyquist criterion can still be estimated accurately. F_{signal} denotes signal frequency and is equivalent to RR. $F_{sampling}$ denotes the sampling frequency and is equivalent to HR. AE denotes the absolute error.

Figure 3.11 shows a scatter plot of all HRs and label RRs in the testing dataset, with each sample color coded by their absolute errors (AE) based on RRs estimated from the fused U-Net. All samples in the testing dataset satisfied the Nyquist criterion. Moreover, we did not observe considerable changes in AEs when the ratios of HR to RR approached the boundary of the criterion.

3.4 Discussions

3.4.1 U-Net of Single Modality

From Figure 3.7, The observed improvements led by U-Net can be attributed to its capability to learn denoising filters and to combine local and global information. On the other hand, the higher MAEs due to the U-Net can be explained by the different levels of the robustness against artifacts of the AM features. For example, we discovered that these poor performing EDRs and SDR did not exhibit a strong coupling with respiration for some subjects, likely due to poorer representative power of the respiratory features. Hence, the

training data can contain outlier segments, which were reflected on the convolutional filters learned. In this case, predictions made for the testing data can be unexpected.

Since the inputs of the U-Net are the ST representations of the DR signals, one can use this data-driven U-Net to denoise other variants of DR signals. For example, DR signals extracted from different physiological signals, feature extraction methods, and study protocols can be conveniently used to train the U-Net if the data are provided. Therefore, it is in nature translatable to data with a different distribution.

In the context of RR estimation, the contractive path is analogous to hierarchical feature extraction. The shallower layer learned the small-scale features in the ST domain. The respiratory information at each time step was correlated with its immediate neighbors. This was expected as breath-by-breath RRs for naturally breathing subjects were unlikely to change drastically during treadmill recovery. We termed this local dependency. Meanwhile, the deeper layer learned the large-scale features in the ST domain. On a coarser level, RRs span across a longer period should also be closely distributed for the same subject. We termed this global dependency. Subsequently, the expansive path up sampled global, coarse information for final prediction. The skip connections provided a mechanism for local, precise information that had not been down sampled by the contractive path to also flow toward the output layer. Together, the expansive path—along with the skip connections—combined both local and global information for final predictions.

3.4.1.1 U-Net of Fusion Modality

The results in Figure 3.7 suggest that the fusion technique should be designed carefully since fusion achieved by simply averaging estimated RRs is not necessarily

superior to single modality. Even if the U-Net was not able to denoise a DR input, it could still utilize partial information in that DR input through fusion because of the complementariness between ECG-R_{AM} and SCG-PT_{AM}. In contrast, the simple fusion method clearly could not take advantage of the multi-modality of the wearable patch.

Certain fusion strategies were based on selecting or assigning weights to RRs estimated from different DR inputs. Such strategies could pose a problem: when all DR inputs have lower SNR, they could hardly be used to estimate accurate RRs. This problem is usually mitigated by discarding data with low SNR, but doing so might lead to significant amount of data wasted when there were better alternatives available. Our fused U-Net performed well without discarding outliers aggressively, though this achievement might be specific to our dataset. One advantage of fusion through U-Net, besides being simple, is that it is very effective at filtering non-respiratory artifacts. Even when all DR inputs to be fused were derived from a segment of low SNR, the U-Net could still utilize temporal dependencies and incorporate information from the preceding and succeeding segments. Finally, since extending the U-Net for fusing different DR inputs simply requires changing the channel size and the training data, it is easily adaptable for different datasets.

3.4.2 The Complementariness of EDR and SDR

We have highlighted the performance of fusion achieved by U-Net, and we will now provide some qualitative explanation to the importance of fusing EDR and SDR. Again, Figure 3.9 shows that non-respiratory artifacts do not necessarily occur simultaneously for different DR inputs. We hypothesize that non-respiratory artifacts in one input of U-Net (e.g., ECG-R_{AM}) can be compensated by a separate input (e.g., SCG-

PT_{AM}). There are various reasons for the observed discrepancies in robustness of the signals against noises. We suspected that the accumulation of sweat during treadmill recovery might affect adhesiveness of ECG electrodes and therefore lower its SNR. Such effect should be minimal for SCG in this situation because SCG is robust against changes in electrical property of the tissue. On the contrary, EDR is more robust against subtle movement. Subtle movement can lead to corrupted SCG measured. Further, SDR might suffer from crosstalk between different modulation sources. For instance, different respiratory phases could alter the characteristics of the spring dashpot system that coupled the subject and the SCG recording system [94], which led to variation in SCG morphology [17], [95]. Simultaneously, the regulation of respiratory centers on the contractility of the heart could also lead to variation in SCG morphology. The two mechanisms may modulate (through AM) the SCG in opposite directions, hence cancelling each other. Unfortunately, it is challenging to measure the extent of the independent contribution to SCG modulation by these two sources. It is beyond the scope of this work to study the mechanisms of SDR, but further investigation could help improve SDR.

3.4.3 Experimental Protocol Extrapolating to Unseen RRs

Results presented in Figure 3.10 were expected from the fused U-Net because the patterns it learned are translational invariant. Specifically, filters trained to remove artifacts at higher RR can be applied to data at lower RR since ST patterns at different spectral bin and temporal step can be similarly distributed. In our cascaded framework, the complexity of data was conveniently abstracted away through the demodulating block. Only simple, translational invariant patterns were observed in the ST representations, which explained the denoising capability of U-Net. Due to random splitting of the training and testing

datasets, we cannot control their RR distributions. Though we did not have examples to show that U-Net can extrapolate denoising for RRs higher than those in the training dataset, we still expect U-Net to work well in such a scenario.

3.4.4 The Effect of Ratio of HR to RR on Performance

Since the extracted EDRs and SDR are sampled at each cardiac cycle, it is important to study how different levels of HR could affect RRs estimated by the fused U-Net. In the feature extraction step of the demodulating block, HR is equivalent to the sampling frequency, and RR is equivalent to the signal frequency. According to Shannon-Nyquist theorem, the sampling frequency should be at least two times the maximum frequency in the signal [96], [97]. Equivalently, the ratio of HR to RR should be greater than two (namely the Nyquist criterion). However, it is uncertain whether uneven sampling can give rise to aliasing within the Nyquist criterion. The illustration in Figure 3.11 suggests that HR-sampled methods (including AM and FM) generally do not suffer from aliasing. Therefore, the BW demodulating scheme, with a high sampling rate at 500 Hz, does not offer substantial advantages in this regard. This finding is consistent with that described in [17] for SDR.

3.4.5 Alternative Respiratory Modulations

In addition to AM, we also explored estimating RR from FM and BW. However, the ST representations were of poor quality in most subjects and thus the U-Net could not learn the expected denoising filters. Though we did observe respiratory sinus arrhythmia from a handful of subjects, FM-based DR inputs did not lead to consistently robust RR estimation. The quality of the BW-based DR inputs also varied among the subjects, most

likely hindered by subtle movements. Meanwhile, due to the discrepancies between the motion artifact frequency and the carrier frequency of both QRS complex and the systolic profile, AM-based DR inputs were less affected by such artifact sources.

3.4.6 Limitations and Future Work

There are a few limitations of the proposed framework. First, transforming the DR signals to ST representations is the most computationally costly step. Though resampling the DR signals to 4 Hz instead of 500 Hz can simplify the computation, STFT may still be impractical for real-time implementation. Therefore, in future work, we may consider replacing STFT by the validated variable frequency complex demodulation (VFCDM) method for its computational efficiency [98]. Note the denoising U-Net generates outputs relatively quickly and is not the computational bottleneck.

Second, we only analyzed the treadmill recovery portion of the dataset, where the presence of severe motion artifacts was minimal. Yet, in the setting of at-home monitoring, subjects could constantly produce motion artifacts that cannot be filtered easily. Hence, unseen edge cases where non-respiratory artifacts cannot be denoised from EDRs and SDR are expected. In these cases, it is likely the cascaded framework cannot achieve accurate RR estimation even with fusion. One example could be the simultaneous and consistent presence of artifacts induced by instability in skin-electrode interface and motion, consequently leading to the low quality of EDR and SDR. Extending the cascaded framework to other portions of the dataset—e.g., treadmill walking and outdoor walking—could help us better understand how the proposed framework works toward these edge cases. Future analyses are required on the remaining portion of the dataset to determine

whether modifications are necessary to the preprocessing pipeline. We believe modifications in the demodulating block, if necessary, such as replacing the Pan-Tomkins algorithm with a wavelet-based method for R-peak detection [99], inclusion of other robust respiratory features such as QRS slopes and R-wave angle [67] especially when motion artifacts commonly seen in normal daily activities are present, and inclusion of other physiological signals such as PPG can together help enable RR estimation for these edge cases without changing the data-driven denoising block—U-Net. Finally, we hypothesize that incorporating an algorithm that detects the resting state of the subject can efficiently translate this work to at-home monitoring.

Third, we only consider a healthy population. Since this is the first time we designed an RR estimation algorithm for this device, it is important to first ensure the validity on subjects without abnormal physiological signals. Although this dataset is sufficient for capturing a wide range of RRs, it does not necessarily reflect the dynamic of physiological states of a pathological population (e.g., patients with pulmonary disease). We would also like to point out that physiologically abnormal ECG and SCG might be expected from the pathological population (e.g., patient may have comorbidities such as cardiac diseases). Albeit we have demonstrated the compatibility of a robust EDR (ECG-SR_{AM})—validated using data from patients with atrial fibrillation (AF) by Kontaxis *et al.* [69]—with our framework on healthy subjects, it is uncertain whether U-Net can denoise ECG-SR_{AM} of AF patients successfully if not trained on these patients. One solution to mitigate such a problem is to validate the cascaded framework on the pathological population (e.g., COVID-19 patients and sleep apnea). Though the dynamic of RRs could

be different, the local and global dependencies modeled by the U-Net might still be partially transferrable to the target population through domain adaption techniques.

These limitations, if successfully addressed, can help demonstrate the translatability of the data-driven U-Net. The inclusion of target data (e.g., during activities other than treadmill recovery or from a pathological population) to the training dataset and modification to the demodulating block will better adapt the U-Net for its designated use.

3.5 Conclusion

In this chapter, we have demonstrated that a chest-worn wearable patch can be used to estimate RR accurately. The advantages of the wearable patch include its suitability for normal daily activities. Also, the proposed U-Net achieved non-respiratory artifacts denoising through temporal dependencies modeling, non-linearity mapping, and multi-modal fusion in a data-driven manner. Although EDRs performed better than SDR, we showed that fusing ECG-R_{AM} and SCG-PT_{AM} through the U-Net resulted in the lowest MAE. We also demonstrated that the proposed framework performed well on unseen RRs, suggesting the generalizability of the cascaded framework. Finally, we concluded that AM-based DR signals were not susceptible to aliasing when the ratios of HR to RR were within the Nyquist criterion. There are a few disadvantages of this work. Since our dataset is consist of a healthy population, it is hard to determine whether the conclusions made were generalizable to the target population. For this reason, the future work will focus on evaluating the proposed method on the target population. In addition, to further improve the cascade framework, it is important to examine the extents to the underlying mechanisms of the respiratory modulation on the systolic profile of the SCG. The results

achieved in this work, together with the future work, will help validate the robustness of the chest-worn wearable patch and algorithms against artifacts towards at-home monitoring of pulmonary diseases.

The example presented in **CHAPTER 2** to replace one functional module of a classical DSP pipeline using a data-drive model has shown to be successful, and this chapter has further reinforced the validity of this approach. We have therefore concluded **Aim 1**. Nevertheless, these two examples only used data-driven models that served one function: feature extraction or denoising. Is doing so the best approach? Why should we limit the potential of data-driven models by setting an arbitrary rule. Thus, we will further examine the possibility to replace more functional modules with ML/DL and perhaps, to leverage novel DL regularization techniques developed in other fields. In the next two chapters, we will proceed to address different accuracy problems in health wearables with this theme.

CHAPTER 4. USING MODALITY-ATTENTIVE FUSION AND U-NET DENSIOING FOR RESPIRATORY RATE ESTIMATION DURING WALKING

RR is an important vital sign to monitor out of the clinic, particularly during physiological challenges such as exercise; unfortunately, ambulatory measurement devices for RR are typically obtrusive or inaccurate. We therefore investigate algorithms that would allow for accurate and robust RR monitoring with a convenient and small chest-worn wearable patch during walking and exercise recovery periods.

In **CHAPTER 2** or **CHAPTER 3**, one functional module of the classical DSP pipeline has been replaced with a data-driven method to improve estimation performance. In this chapter, we expanded this strategy by replacing two functional modules in a classical DSP pipeline to estimate RR more accurately during walking: fusion and denoising. When compared to **CHAPTER 3**, the chest-worn wearable patch has also received an upgrade with a new sensing modality: PPG. The introduction of the new modality has brought an immediate research opportunity—demonstrating improved RR estimation accuracy by fusing ECG, SCG, and PPG. After carefully examining EDR, SDR, and PDR signals, we found that these signals can be intermittently contaminated by motion artifacts. Such finding consequently motivated the development of a new fusion method, which attempted to adaptively weigh the modalities based on the signal quality. The increased complexity of the accuracy problem drove the development of a DL model that combined a modality-attentive fusion module with the denoising U-Net proposed in **CHAPTER 3**. Surprisingly, since the fusion module developed was analogous to the DSP-based RQI_{FFT}, we found that the fusion weights (or attention weights) were interpretable and informing towards the

quality of the modalities. Ultimately, such finding forecasted our interest to study interpretable DL models in later chapters.

4.1 Introduction

RR is an important vital sign to track since abnormal RR can be indicative of underlying cardiovascular or respiratory conditions. For example, unexpected changes in RR may be due to lowered efficiency of oxygen delivery to tissue (e.g., pneumonia) or adaptation to pathological anatomy of the respiratory systems (e.g., chronic obstructive pulmonary disease [COPD]) [100]. Besides respiratory diseases, acute cardiovascular events may also have respiratory origin. It has been shown that tracking RR can help predict cardiac arrest (CA) as respiratory deterioration is also one common cause of in-hospital CA [101]. Nevertheless, RR is not typically monitored for persons outside of clinical settings.

In particular, the reason for the lack of RR measurements at home is likely because the measurement devices can be cumbersome to wear or obtrusive for ADLs. Although HR is easier to measure, HR can be modulated for various reasons that are unrelated to the target conditions. Consequently, relying on HR alone may lead to reduced specificity for prediction or detection of target conditions, which is especially detrimental for life-threatening events such as cardiac arrests. Unfortunately, about 90% of the patients who experienced Out of Hospital Cardiac Arrests (OHCAs) died. Moreover, 70% of these OHCAs occur at home. Among the Emergency Medical Services-treated OHCAs, half of them are unwitnessed, which yields a survival to hospital discharge rate of a shockingly low 4.7% [102]. Measuring RR at home for patients with CA risk is not a common practice,

but its predictability for CA [103] suggests it should be. Moreover, with the recent growth in remote patient monitoring solutions, a device with algorithms enabling accurate RR monitoring for patients in everyday settings would be of great value.

4.1.1 Robust Respiratory Monitoring using Wearable Devices

To fulfill its predictive power for sudden pathological events in the target population, at-home RR monitoring devices should be both unobtrusive to ADLs and sufficiently accurate in realistic scenarios when compared to gold standard devices (e.g., nasal canula or a face mask that analyzes gas flow). A miniaturized chest-worn wearable patch is an attractive form factor since it can measure respiratory modulated ECG, SCG, and PPG signals continuously and non-invasively, is unobtrusive to ADLs, and reduces stigmatization risk. In fact, this form factor appears well suited for a diverse range of respiratory monitoring goals (e.g., adverse cardiac events, pneumonia, clinical deterioration, pain, cognitive load, and environment-induced stress) highlighted by [104].

Besides the form factor, the synergistic nature of the wearable patch also makes it an ideal choice for respiratory monitoring. Previously, this wearable patch has also demonstrated accurate estimation of several cardiopulmonary parameters such as RR [18], oxygen uptake [35], stroke volume [15], [16], peripheral oxygen saturation [105], heart rate (HR) [106], and tidal volume [107]. Besides health parameter estimation, it has also demonstrated the ability to classify heart failure status [33] and physical activities [108], [109], which can provide contextual information that is usually overlooked due to the difficulty to collect these types of data. RR estimated using this wearable patch and a denoising U-Net [18] was assessed during exercise recovery with minimal motion, which

demonstrated the feasibility estimate a wide range of RR accurately. However, the at-home population may move considerably. Thus, it remains uncertain whether such a device can be used to predict pathological events in at-home settings.

4.1.2 Heterogeneous Respiratory Quality During Walking

In at-home setting, it is critical to first demonstrate accurate RR estimation during walking for two reasons: 1) walking is frequently used as a cardiopulmonary test to assess disease severity and 2) it is a harder task for RR estimation when compared to other at-home activities such as watching TV, working, cooking, etc. because of motion artifacts. Indeed, motion artifacts can corrupt the multi-channel, multi-modal physiological signals measured by this wearable patch to different extents—namely heterogeneous respiratory quality. This makes estimating RR from these signals jointly a challenging problem. Note that here we refer to the modality as the sensor type (e.g., ECG, PPG, or ACC sensors), and we refer to the channel as the unique measurements within the same modality (e.g., different wavelengths for PPG sensors or different directions for ACC sensors).

4.1.3 Demodulate Respiration from Physiological Signals

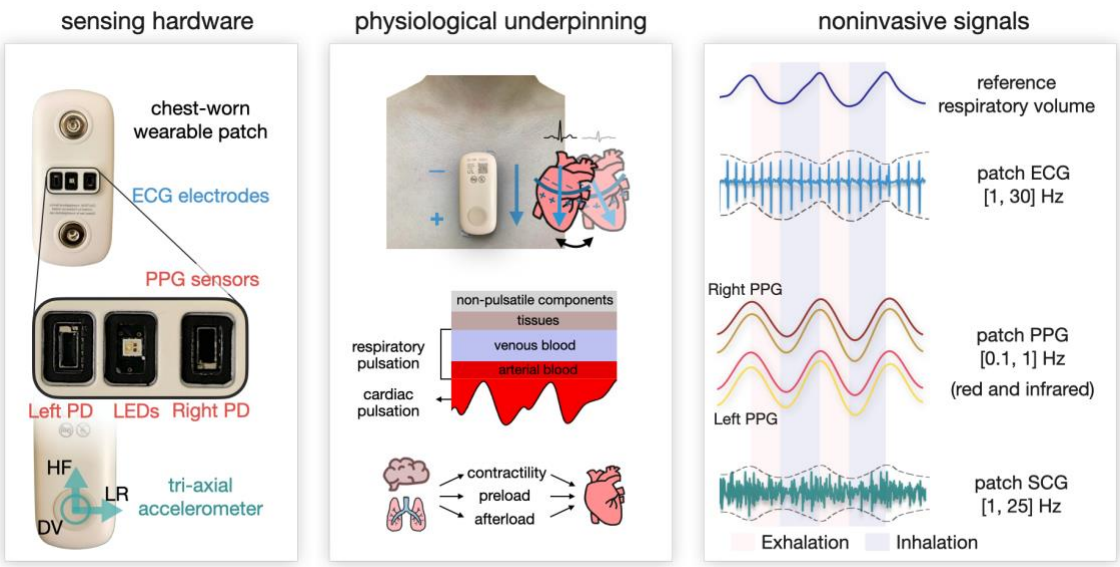


Figure 4.1. Respiratory modulation induces variations in cardiac axis dipole, contractility, preload, afterload, arterial blood volume, and venous blood volume, which are manifested in ECG, SCG, and PPG signals. DV: dorsal-ventral, HF: head-to-foot, LR: left-right, PD: photodiode, LED: light emitting diode.

The heterogeneous respiratory quality can be attributed to the differences in respiratory modulation principles of the physiological signals. As shown in Figure 4.1, The extraction of respiratory signals from a non-invasive wearable device is based on AM, BW, and FM [20]. The AM of ECG signal originates from the cardiac axis rotation due to respiratory movement [13], [23]. The tri-axial ACC signals can be modulated in two ways primarily. First, respiration can modulate the contractility, preload, and afterload of the heart, which results in AM of the SCG signals [17]. Note that SCG measures the chest wall vibrations in response to cardiac contraction. The BW of ACCs signal can also capture the chest wall movement due to respiration. Likewise, PPG signals are also modulated primarily through two mechanisms. First, the cardiac pulsation of the arterial blood leads to AM of PPG since arterial blood can be modulated by respiration-induced changes of stroke volume and autonomic mediation of the arterial smooth muscle [20]. Second, BW of PPG signals is due to the respiration-induced changes in intrathoracic pressure, which

consequently modulate both the arterial [20] and the venous blood [110]. Notably, such changes are less affected by body position or movement [111]. Respiratory movement may also contribute to the BW components in PPG signals. However, we found that the relative amplitudes of the multi-wavelength PPG signals at the respiratory frequency have varying strengths, suggesting the BW components mainly originate from the variations in blood volume tissues instead of motion [112], [113]; the intensity of motion-induced variations is usually equally presented in PPG channels.

Because respiration manifests itself through different mechanisms in the channels and the modalities, we can assume the sources of errors (e.g., motion artifacts) may also manifest themselves in these signals to different extents. In this work, we hypothesize that by adaptively selecting and fusing these signals depending on respiratory quality, we can better address the heterogeneous respiratory quality of the measured signals during walking. The contributions of this work are: 1) a novel modality-attentive (MA) fusion method to combine multi-modal data, 2) an RR estimation pipeline that combines ECG and chest PPG to estimate RR during walking, 3) demonstration of the cross-modality generalizability of the proposed pipeline, and 4) validation of the compatibility of a widely used respiratory quality index (RQI) with our pipeline.

4.2 Methods

4.2.1 Datasets

Table 4.1 The source dataset (D_{source}) and the target dataset (D_{target}).

	D_{source}	D_{target}
Modalities and channels	ECG x 1 ACC x 3: DV, HF, and LR (from wearable patch [30])	ECG x 1 ACC x 3: DV, HF, and LR PPG x 6: red, infrared, and green on L and R sides (from wearable patch [100])
Ground truth	RR (from COSMED K5 mask)	RR (from COSMED K5 mask)
Activities	treadmill recovery	sitting, 6MWT, 6MWT-R, self-paced walking, self-paced running, exercise recovery
N_{sub}	17 (10M, 7F)	17 (11M, 7F)
Weight [kg]	67.5±14.1	68.3±17.8
Height [kg]	170.5±9.9	173.8±12.7
Age [years]	26.8±4.1	24.2±2.4

As can be seen in Table 4.1, two datasets were used in this work: the Source Dataset (D_{source}) and the Target Dataset (D_{target}). D_{source} was only used for training, and D_{target} was used for training and testing. These datasets have been collected using protocols approved by the Georgia Institute of Technology Institutional Review Board (H18451 and H21100 respectively). The datasets have been de-identified and are used here for post-hoc analysis only. The ground truth RR of these two datasets was measured using the COSMED K5 wearable metabolic system (COSMED, Rome, Italy), sampled at each breath. This device reported limits of agreement of 0.30 ± 2.67 breaths per minute (brpm) for their RR measurements [88]. The physiological signals were measured by a small wearable patch with a dimension of 79 mm × 36 mm × 15 mm, attached to the mid-sternum of the subjects through ECG electrodes. Although the wearable patches for both datasets have similar form

factor, the devices used in these two datasets differed in the number of modalities collected and the specific procedures of the study protocol.

D_{source} was previously presented and described in detail by Shandhi *et al.* [35]. This dataset includes 17 healthy subjects. The wearable patch measured two modalities: ECG and tri-axial ACC signals. The activities performed by the subjects include baseline measurement in a seated posture, treadmill walking, treadmill recovery in a standing posture (five minutes), outdoor walking, and walking recovery in a standing posture. Only data from the treadmill recovery activity was used in this work. The ground truth RR collected during this period ranged from 5-40 brpm (equivalently 0.083-0.67 Hz).

For D_{target}, 21 healthy subjects were recruited, and 17 of these subjects were analyzed in this work. Three subjects were excluded due to unsatisfactory ECG quality. Since the protocol involved exercise, sweat reduced the adhesiveness of the ECG electrodes used for these subjects and therefore deteriorated the ECG quality. After initially observing this problem, ECG electrodes with added adhesiveness (3M™ Red Dot 2560 ECG electrodes, St. Paul, Minnesota) designed for stress testing were used instead. Another subject was excluded due to occasional speaking, which may lead to errors in the ground truth RR. The chest-worn wearable patch used in this study was described in prior work [105], and is an updated version of the wearable patch used in Shandhi *et al.* [35]. Besides ECG and tri-axial ACC, this updated wearable patch also measured an additional modality: two sets of multiwavelength (red [660 nm], infrared [950 nm], and green [526 nm]) PPG signals from the left (L) and the right (R) sides of the midsternal line. Of the seventeen subjects, three subjects still exhibited noisy ECG despite the usage of the more adhesive electrodes, likely due to faulty fabrication of the ECG snap buttons. Since we also placed

two additional wearable patches near the clavicle bone and on the left side of abdomen, we mitigated this by using the ECG from either of these two locations instead (selected based on visual inspection). We always used ACC and PPG signals from the mid-sternum patch. In both datasets, the ECG signals were sampled at 500 Hz, and the ACC signals were sampled at 1000 Hz. The PPG signals in D_{target} were sampled at 67 Hz.

The study protocol of D_{target} included a series of tasks, starting with stationary measurements in several postures: upright (three minutes), supine (one minute), laying on the left (one minute), laying on the right (one minute), and prone (one minute). Next, subjects conducted the Six-Minute Walk Test (6MWT) [114] and 6MWT with resistive breathing (6MWT-R: noses were clamped and subjects were asked to breathe through a straw with a diameter of six mm) in a randomized order. Note that using one straw can mimic the Global Initiative for Chronic Obstructive Lung Disease (GOLD) stage four severity in healthy subjects [115]. For both 6MWT tasks, subjects were instructed to walk as fast as possible without feeling uncomfortable. After performing these two variants of the 6MWT, subjects conducted three self-paced activities: stair climbing (three floors down, four floors up, and finally one floor down), outdoor walking (in a predetermined route that includes a flat surface, an inclined surface, a declined surface, and stairs), and outdoor running (one lap of a standard track field, 400 m). Following each physical task (i.e., 6MWT, 6MWT-R, stair climbing, outdoor walking, and outdoor running), subjects sat in an upright posture for three minutes, to capture physiological changes during exercise recovery.

In this chapter's analysis, we grouped the activities used in this work into three types: stationary activities (stationary measurements in an upright posture and all exercise

recovery periods), walking (6MWT, 6MWT-R, and self-paced outdoor walking), and running (self-paced outdoor running). Other activities were excluded from this analysis due to their durations not being sufficiently long (< two minutes) to apply the denoising deep learning model for some subjects. The ground truth RR collected in D_{target} ranged from 5-50 brpm (equivalently 0.083-0.83 Hz).

4.2.2 System Overview

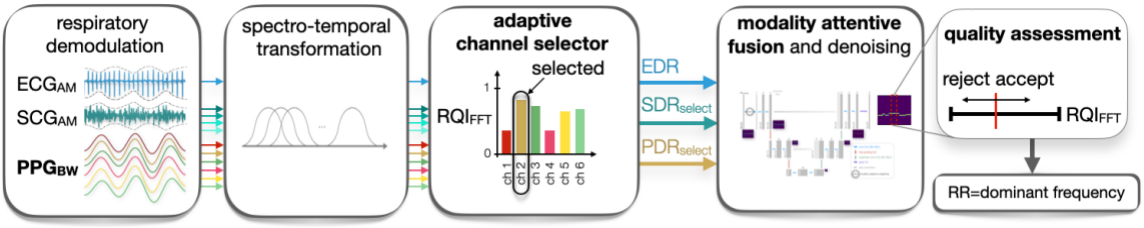


Figure 4.2. System overview of a RR estimation pipeline. Physiological signals were first demodulated, transformed to the ST domain, selected for the optimal channel, combined through modality-attentive (MA) fusion, denoised using U-Net, and rejected if the quality was low. RR was finally estimated from the denoised data. AM: amplitude modulation. BW: baseline wander. RQI_{FFT}: respiratory quality index based on Fast Fourier Transform.

As presented in Figure 4.2, estimating RR from ECG, PPG, and SCG involved several steps. First, we applied signal demodulation techniques to extract DR signals, which were next transformed to the ST domain. To reduce the complexity of the multi-channel data, an adaptive channel selector was used to select one channel for a modality. Next, a proposed fusion method—MA fusion—was used to adaptively merge the multi-modal data, followed by U-Net-based deep learning model [18] to denoise the fused data. The estimated RR was then determined from the denoised spectrogram. Lastly, to improve the transparency in the quality of RR estimated from this wearable patch, we used a popular

choice of respiratory quality index (RQI) to identify outliers and improve RR estimation accuracy.

4.2.2.1 Respiratory Demodulation of Physiological Signals

To derive DR signals from ECG, we first extracted the slope range features introduced by Kontaxis *et al.* [69]. The extracted features were interpolated, smoothed using a fourth order Butterworth filter with a bandwidth of [0.08, 0.8] Hz, and downsampled to 5 Hz. This EDR signal is denoted as EDR_{SR}. DR signals derived from ACC signals were first smoothed using a fourth order Butterworth filter with a bandwidth of [1, 25] Hz to extract tri-axial SCG signals. Next, the smoothed peak SCG features were extracted using a method described in [18] from different channels: dorsal-ventral (DV) axis, head-to-foot (HF) axis, left-right (LR) axis, and their ℓ^2-norm. Again, these features were interpolated, smoothed, and downsampled to 5Hz. These four derived SDR signals are denoted as SCG_{DV}, SCG_{HF}, SCG_{LR}, and SCG_{norm} respectively. Finally, the PDR signals were extracted by first smoothing the raw PPG signals using a fourth order Butterworth filter with a bandwidth of [0.08, 1] Hz and next downsampling them to 5 Hz. Four PDR signals were extracted, and they are denoted as $PPG_{red,L}$, $PPG_{red,R}$, $PPG_{infrared,L}$, and $PPG_{infrared,R}$. The EDR and SDR are based on the AM principle while the PDR signals are derived based on the BW principle. The passbands of for processing ECG and SCG were determined based on [18]. The passband for processing PPG signals were determined using the physiological plausible RR range converted to Hz.

There were other DR signals that could be derived, but they were not included in this work because they showed compromised respiratory quality. $PPG_{green,L}$ and $PPG_{green,R}$

were excluded because they were consistently less accurate when compared to PPG_{red} and $PPG_{infrared}$ signals for RR estimation. BW-based DR signals derived from ECG and ACC signals were more easily corrupted by motion artifacts during walking and therefore were not used in this work. The AM-based PDR signals also showed low signal quality, likely due to the poorly perfused sternal skin. The FM-based DR signals were excluded because they were not consistently reliable across subjects.

4.2.3 Spectro-Temporal Transformation

The DR signals were first normalized using a method based on a median filter [18] and next transformed to the ST domain via STFT. We used a window size of 60 seconds with 95% overlap. The resulting spectrogram is truncated to a spectral dimension of 0.03 to 0.98 Hz to retain only RR-relevant information and to reduce computational complexity. Each spectrogram has 20 temporal steps with a temporal resolution of three seconds, using 55 frequency bins with a spectral resolution of 0.0167 Hz (equivalently, 1 brpm). We will refer to the spectrograms as data or DR data hereinafter.

4.2.4 Adaptive Channel Selector

Some channels can be more representative of respiration than others. For example, we observed that PPG_{red} and $PPG_{infrared}$ are generally better than PPG_{green} for RR estimation. Doing so could help reduce the complexity of the multi-channel data (from 11 DR signals to 3 DR signals). The Fast Fourier Transform (FFT)-based RQI (RQI_{FFT}) that has been used in [78], [116], [25], [117], [118] was adopted here because Khreis *et al. [116]* has shown that RQI_{FFT} performs better compared to the alternative RQIs for automatic selection of the DR signals during walking. RQI_{FFT} can be computed as,

$$RQI_{FFT} = F_{max} / F_{total} \quad \textbf{(4.1)}$$

where F_{max} was the maximum of the sum of two consecutive power spectral densities, and F_{total} was the sum of total power spectral densities within the respiration range (0.03 to 0.98 Hz).

In this work, the middle two temporal steps of the spectrogram were first averaged to generate a spectral vector. This vector was next used to compute RQI_{FFT}, which ranged from 0 to 1. The same operation was repeated for the spectrogram of each SDR and PDR channel, and the optimal channels for each modality (defined as having the highest RQI_{FFT}) were selected and denoted as SCG_{select} and PPG_{select} as illustrated in Figure 4.2.

4.2.5 Modality-Attentive Fusion with U-Net Denoising Block

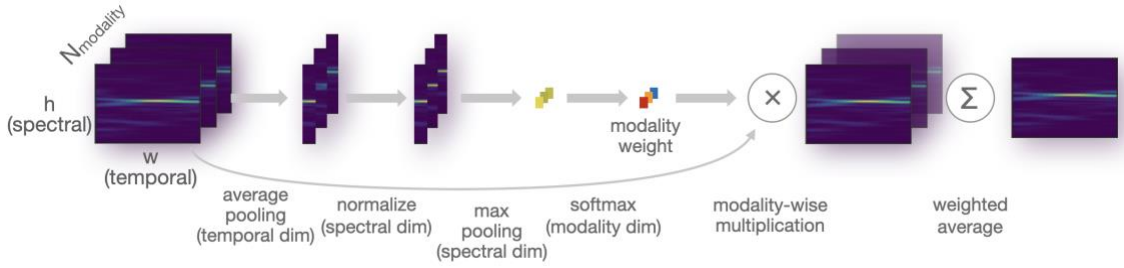


Figure 4.3. Modality-attentive (MA) fusion. Average pooling was first applied to the temporal dimension of the multi-modal data ($N_{modality}$ x h x w). The $N_{modalitiy}$ condensed vectors were next normalized, so each vector summed to one. Max pooling was applied on each vector to further reduce the dimension of the data to $N_{modalitiy}$ x 1. To generate the modality weights, softmax layer was applied. Finally, we multiplied the multi-modal data by the modality weights using modality-wise multiplication, and the weighted average was computed to generate the fused ST data.

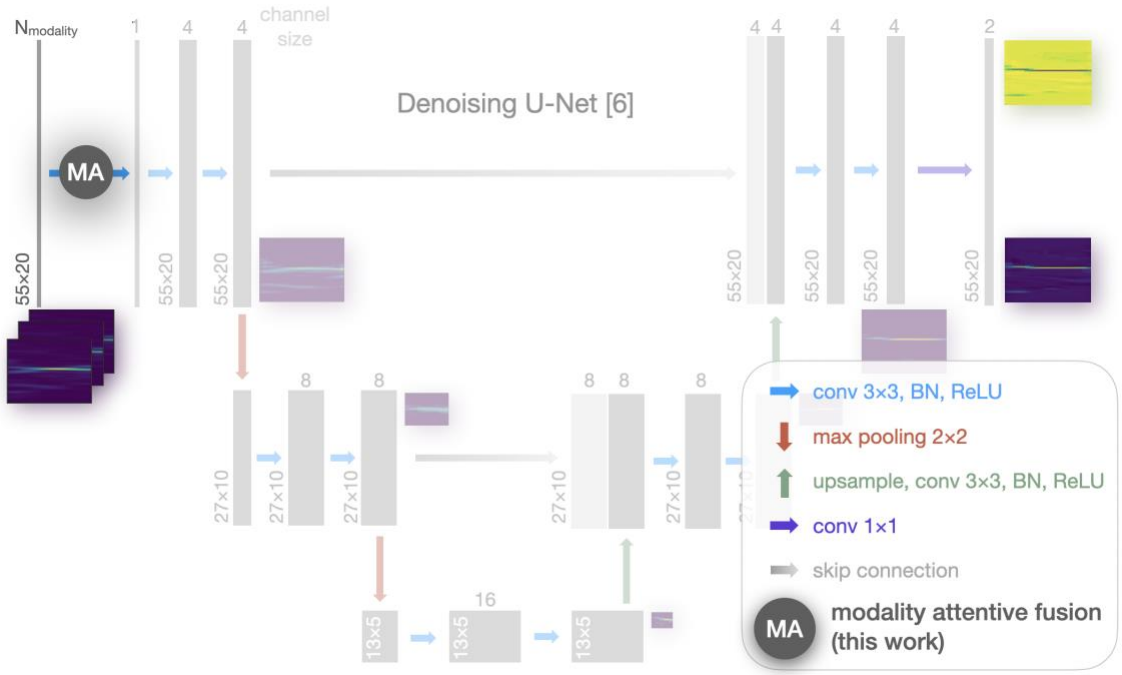


Figure 4.4. MA-U-Net. This architecture was used to denoise ST representation of the wearable patch data. MA fusion was applied at the first layer to reduce the data from $N_{modality}$ modalities to one. Some of the outputs of the layers were also shown.

A U-Net denoising block for reducing non-respiratory noise was previously proposed in [18] to achieve ST modeling and modality fusion. Fusion through this method was accomplished using the convolutional filters, which applied modality-specific filters on the spectrogram of each modality and summed the filtered spectrograms. During inference, the parameters of these filters were fixed and therefore were not adaptable to unexpected scenarios, especially when respiratory qualities were heterogeneous across modalities. To address this, we proposed MA fusion, which was applied prior to the convolutional layers and fused multi-modal data into a single spectrogram. As illustrated in Figure 4.3, MA fusion was achieved by first using an average pooling layer along the temporal dimension of the multi-modal spectrograms. The averaged spectral vectors were normalized so they each summed to one. Next, max pooling was applied to select the most

prominent spectral feature for each modality, and a softmax layer was applied along the modality dimension to compute the modality weights. The multi-modal data were fused by the modality weights through modality-wise multiplication. Finally, the fused spectrogram was used as the input to the U-Net denoising block as shown in Figure 4.4. This method incorporated the essence of RQI_{FFT} and signal processing-based fusion methods used for HR estimation [119]—the more centered the spectral engergy is around the dominant energy for a measured signal, the more representative such signal is of the signal of interest (e.g., respiratory pulsation). The method that combined MA fusion and U-Net was denoted as MA-U-Net.

The U-Net denoising block was implemented with the same hyperparameters and training procedures (e.g., learning rate, mini-batch size, data augmentation techniques, etc.) used in [18]. The spectrogram contained slightly more frequency bins since this dataset included a wider range of RR. The loss functions were the same as those used in [18].

4.2.6 Respiratory Quality Assessment

One common limitation of wearable devices was the lack of transparency for data quality as suggested by Altini *et al.* [120]. It is critical to provide RQI while estimating health parameters because even validated devices could be inaccurate due to various reasons such as misplacement [104]. In this work, RQI_{FFT} was again used to identify outliers in the denoised spectrogram.

4.2.7 Experimental Design and Evaluation

We described three experiments to study various topics: Sensing Modalities Comparison, Fusion Methods Comparison, and Generalizability of MA-U-Net. Throughout this work, LOSO cross validation was used on D_{target} to ensure generalizability to unseen subjects, and the average and standard deviation of the MAEs across the subjects in D_{target} were reported. Statistical significance was tested using either pairwise t-tests for normally distributed data or Wilcoxon signed-rank tests for nonnormally distributed data (tested using the Shapiro-Wilk test).

4.2.7.1 Experiment 1: Fusion Methods Comparison

In this experiment, five fusion methods were implemented: spectrogram averaging (RR is estimated from the average of the multi-modal spectrograms [121]), RR estimates averaging (averaging RR estimated from modality-specific U-Nets), early-U-Net (early fusion at the first convolutional layer as in [18]), SE-U-Net (after early fusion, a squeeze-and-excitation [SE] layer [122]–[124] was used to highlight salient features for the task), and MA-U-Net (the proposed method). These models were assessed on the three types of activities separately. No outliers were rejected in this analysis.

4.2.7.2 Experiment 2: Sensing Modalities Comparison

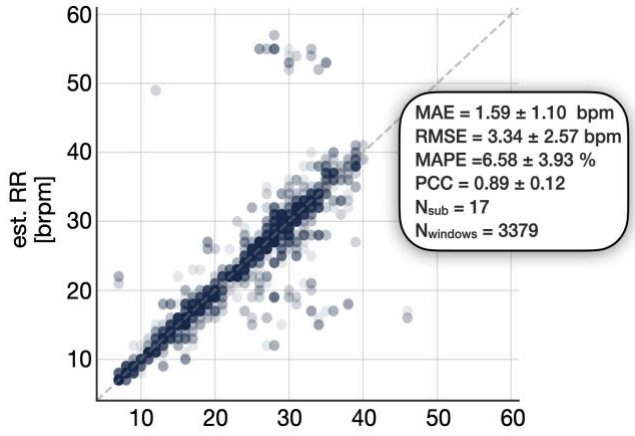
In this experiment, we assessed the accuracies of seven combination of sensing modalities (i.e., *EDR, PDR, SDR, EDR & PDR, EDR & SDR, PDR & SDR*, and *All*). The *EDR* used was ECG_{SR}. The *SDR* used was SCG_{select}. The *PDR* used was PPG_{select}. The fusion and denoising method used was MA-U-Net. Note that for single modality data, applying MA-U-Net was equivalent to applying U-Net directly since the modality weight

was always one for single-modal data. These models were assessed on three types of activities separately. No outliers were rejected in this analysis.

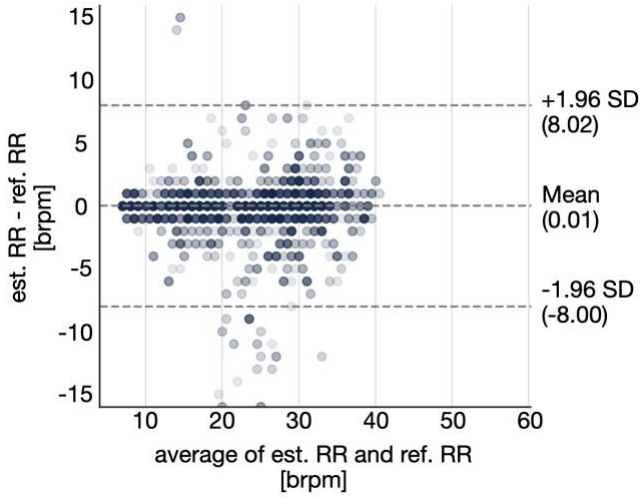
4.2.7.3 Experiment 3: Generalizability of MA-U-Net

To test the generalizability of the denoising U-Net, we trained MA-U-Net using D_{source} and tested the model performance on D_{target}. Specifically, we only trained the model using ECG_{SR} from treadmill recovery periods of D_{source}. MA fusion did not include learnable parameters, and the modality weights only depend on the data. As consequence, the MA-U-Net trained could adapt to a different dataset (e.g., one that has an unseen modality such as PDR, or one that has different number of modalities). We only tested the model on the EDR data and the PDR data of D_{target}. We also experimented with the model trained using SDR data from D_{source}, but the model was less generalizable. This experiment was conducted on the walking activities only. No outliers were rejected in this analysis.

4.3 Results



(a)



(b)

Figure 4.5. RR estimation performance of EDR and PDR fused and denoised by MA-U-Net during walking after rejecting outliers using an RQI_{FFT} threshold of 0.83 (coverage=84.43%). (a) Regression analysis. (b) Bland-Altman analysis.

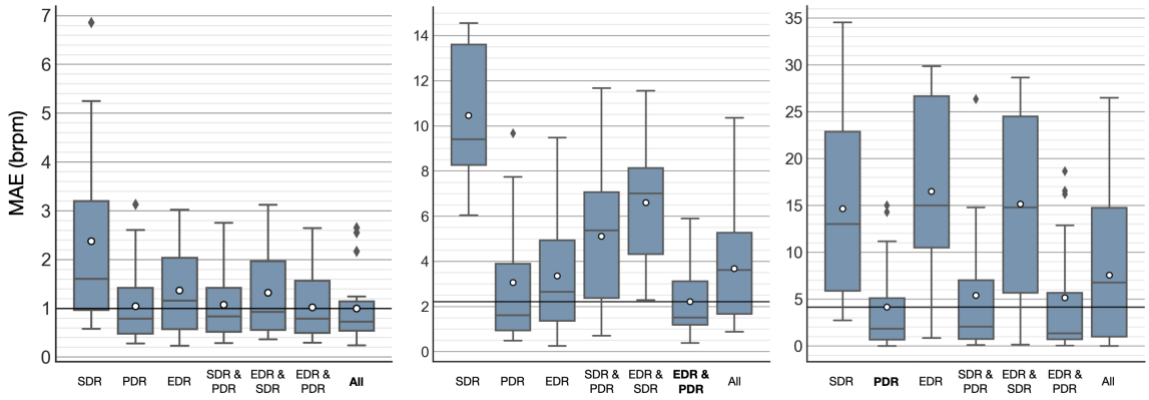
In Figure 4.5, we showed the regression and the Bland-Altman plots for *EDR & PDR* data that were fused and denoised using MA-U-Net and with outliers rejected using an RQI_{FFT} threshold of 0.83 to achieve a coverage near 85%. The MAE, RMSE, MAPE,

and PCC were 1.59 brpm, 3.34 brpm, 6.58 %, and 0.89 respectively. A limit of agreement (LoA) of -8.00 to 8.02 brpm was achieved, with a bias of 0.01 brpm.

Table 4.2 Comparison of RR estimation fusion methods for combining EDR and PDR data.

	spectrogram averaging [116]	RR estimates averaging	early-U-Net [13]	SE-U-Net [118]	MA-U-Net (this work)
stationary	1.32±0.82	1.09±0.75	0.98±0.72	**0.93±0.75**	1.02±0.69
walking	3.75 ± 2.40	2.76 ± 1.65	2.41 ± 1.80	2.32 ± 2.07	**2.21 ± 1.66**
running	14.97±8.88	8.80±5.43	5.57±6.91	**4.44±5.68**	5.14±6.53

Table 4.2 reported the results for Experiment 1. During walking, the MAE was the lowest when using MA-U-Net (2.21 brpm). During stationary activities, the MAE was the lowest when using SE-U-Net (0.93 brpm). During running, the MAE was the lowest when using SE-U-Net (4.44 brpm), though we noticed that using the PDR data alone was more accurate than fusing EDR and PDR data regardless of the fusion method for the running periods.



(a) stationary activities (b) walking (c) running

Figure 4.6. Comparing mean absolute errors (MAEs) of different sensing modalities using MA-U-Net during (a) stationary activities. (b) walking. (c) running. Note that the white dots indicate the mean MAE of a modality, and each black horizontal line represent the modality with lowest mean MAE in each activity. The boxplots show the 25th and 75th percentiles and the median value

Figure 4.6 showed the results for Experiment 2. During stationary activities, MAEs for most modalities were lower than 2 brpm except for *SDR*. The three modalities with the lowest MAEs were *PDR*, *EDR & PDR*, and *All* (1.05, 1.02, and 1.00 brpm). During walking, the two modalities with the lowest MAEs were *PDR* and *EDR & PDR* (3.06 and 2.21 brpm). During running, the two modalities with the lowest MAEs were *PDR* and *EDR & PDR* (4.14 and 5.14 brpm).

Table 4.3 Generalizability of MA-U-Net that fused ECG and PPG during walking.

	no U-Net	trained on D_{source}	trained on D_{target}
MAE	3.75 ± 2.40	2.41±1.99*	2.21 ± 1.66*

Note that the MAE is presented in mean ± standard deviation brpm across subjects.
*denotes statistical significance in the pairwise test relative to no U-Net ($p < 0.005$).

Table 4.3 showed the results for Experiment 3. Training on D_{target} achieved the lowest MAE. Surprisingly, when MA-U-Net was trained on D_{source} (using only EDR data

since PDR data is not available in D_{source}), it showed that it could still denoise fused spectrogram of *EDR & PDR*. Its performance was comparable to the model trained on D_{target}.

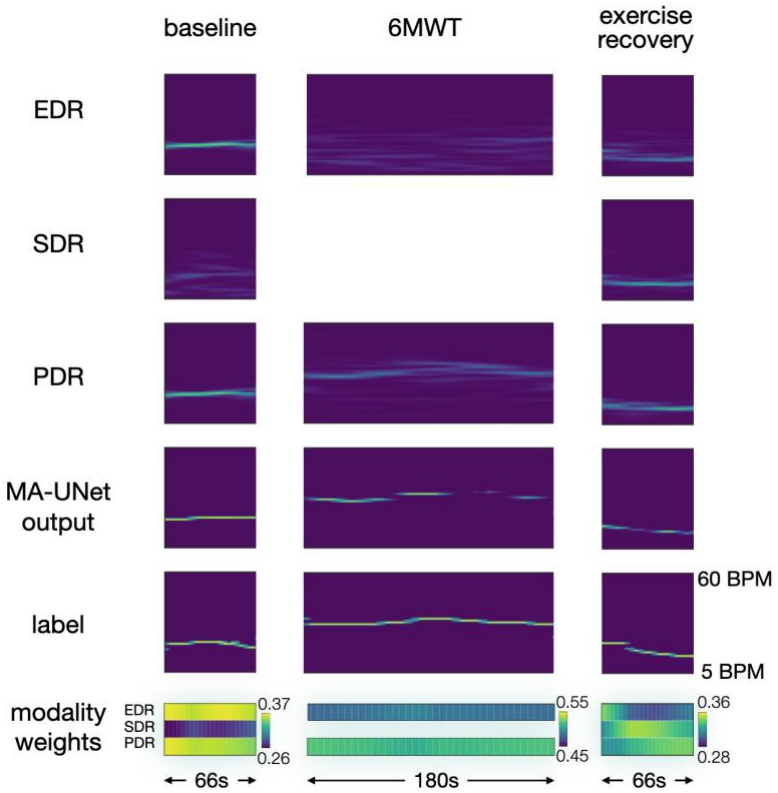


Figure 4.7. Inputs of all modalities, MA-U-Net output, label, and modality weights for three different subjects. During baseline, the EDR and PDR exhibited better signal quality than SDR has. During 6MWT, PDR exhibited better signal quality than EDR did. During exercise recovery, SDR and PDR exhibited slightly better signal quality than EDR did. These observations on the inputs were consistent with the attention weights.

In Figure 4.7, we further provided explainable, qualitative results to demonstrate that MA fusion worked as expected. In fact, the modality weights computed as the intermediate by-products predicted the qualities of spectrograms well. During baseline measurement, the SDR spectrogram for an example subject had poorer respiratory quality whereas the subject's EDR and PDR spectrograms showed better quality. As desired, the weights for the SDR spectrogram were also lower than those of the EDR and PDR

spectrograms. During the 6MWT, the PDR spectrogram was more robust for another subject, and its weights were higher than that of the EDR spectrogram. Finally, during exercise recovery of the last example subject, the weights of the EDR spectrogram reflected its lower quality.

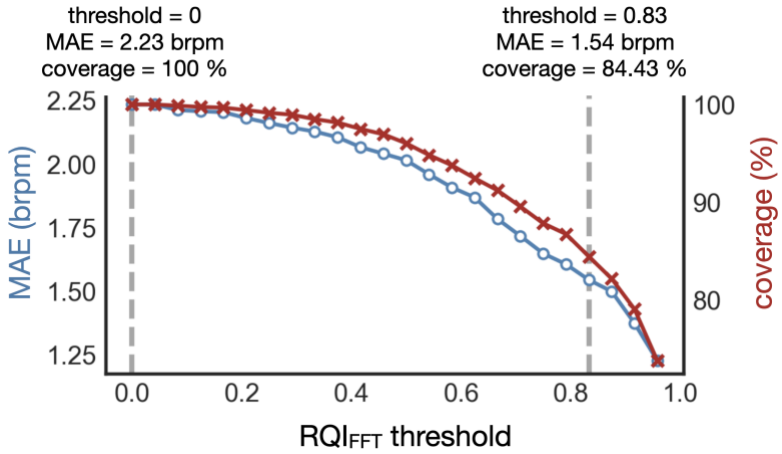


Figure 4.8. Coverage (percentage of samples retained, in %) and MAE analysis using EDR and PDR fused and denoised by MA-U-Net during 6MWT. The x-axis indicated the RQI_{FFT} threshold used to reject the data. The left y-axis indicated the MAE on remaining data from all subjects. The right y-axis indicated the coverage on remaining data from all subjects.

Figure 4.8 studied the relationship between MAE and data coverage (defined as the percentage of samples left after rejecting outliers) as the threshold of the RQI_{FFT}-based outlier rejection algorithm was gradually increased. MAE was monotonically decreasing when higher thresholds for RQI_{FFT} were used, suggesting RQI_{FFT} could effectively identify outliers from the fused, denoised spectrograms.

Table 4.4 State-of-the-art RR estimation methods during walking.

Paper	Sensing modality	Placement	Walking Protocol	N	Performance
Lepine *et al.* [122]	ECG and ACC	Embedded in T-shirt on the chest	Walking on treadmill	8	RMSE = 5.97 brpm (on 30s windows)
Jeyhani *et al.* [120]	IP	Electrodes on the chest	Walking on treadmill (3 km/h and 6 km/h)	15	MAE = 0.35-1.53 brpm (on 15s windows)
Liaqat *et al.* [123]	IMU	Wrist-worn smartwatch	Walking and 6MWT	14	MAE = 0.6 – 1.6 brpm (on 30s window, with 10-20% coverage)
Massaroni *et al.* [124]	Piezoresistive sensors	Embedded in elastic bands were positioned at the trunk	Walking on treadmill (1.6-6.6 km/h)	10	MAE = 1.25 – 1.86 brpm (on 60s windows)
Khreis *et al.* [111]	ECG	Strap around the chest (Zephyr)	Walking for 6 minutes	10	Median of AE = 0.3 - 1.4 brpm (on 32s windows)
Qiu *et al.* [121]	IP	Electrodes on the chest	Walking on treadmill (2.5 km/h and 6 km/h)	15	MAE = 0.17 brpm (breath-by-breath, controlled breathing pace at 20 brpm)
Berkebile *et al.* [113]	IP	Electrodes on the chest	Walking on treadmill (3.21 km/h)	13	MAE = 0.115 ± 0.120 brpm (breath-by-breath)
This work	ECG, sternal PPG, SCG	Electrodes on the chest (smallest footprint)	6MWT, 6MWT with resistive breathing, and self-paced walking	17	MAE = 1.59 ± 1.10 brpm (on 60s windows, denoised with neighboring windows)

As shown in Table 4.4, we compared our performance with other papers studying RR estimation during walking in existing literature, including those using impedance pneumography [125], [118], [126], ECG [116], inertial measurement unit [127], [128], piezoresistive sensors [129]. We only selected papers that examined estimation performance on at least eight subjects during walking since it was the emphasis of this work. These papers assessed their methods using different datasets and thus was not directly comparable with ours. Nevertheless, our results still demonstrated a competitive accuracy while tested on the most subjects with a less controlled walking protocol. Most importantly, the sensing hardware has the smallest footprint. These advantages positioned

this work well for further investigation on usability, utility, and human factors with a goal to drive adoption in clinical setting.

4.4 Discussions

4.4.1 Modality-Attentive Fusion During Walking

The findings of Experiment 1 suggested that during walking, MA fusion best addressed heterogeneous respiratory quality since adaptive method (e.g., MA-U-Net) showed higher RR estimation accuracy than the nonadaptive alternatives (e.g., early U-Net). Further, fusing EDR and PDR data (i.e., spectrogram averaging) performed worse than using PDR data alone when spectrogram averaging fusion was used, suggesting multi-modal data complemented each other only if the right fusion method was used.

Given the scope of this work, walking was the activity that held the most pertinence to at-home RR monitoring using this chest-worn wearable patch. During walking, the lowest MAE was achieved using MA-U-Net on *EDR & PDR* spectrogram, and this can be attributed to MA-U-Net's ability to adaptively fuse multi-modal data.

4.4.2 RR Estimation Accuracy of Different Modalities During Different Activities

Based on the results of Experiment 2, fusing EDR, PDR, and SDR data (*All*) achieved the best performance when motion artifacts were not present. On the other hand, SDR data were the least accurate and could even reduce accuracy when fused with other DR data during walking and running. This was expected since ACC sensors directly measured motion-related information, which included both respiration-modulated cardiac

vibrations and the motion artifacts. Therefore, we suggested fusing SDR signals with other modalities only during stationary activities.

During walking, the EDR data was noticeably less accurate than the PDR data, but it could still provide complementary information if appropriate fusion methods were applied. During running, however, the EDR data was not only the least accurate but also unhelpful when fused with the PDR data. Hence, only PDR data should be used during running.

There could be two possible explanations for the differences in estimation accuracies between EDR and PDR data during different activities. First, although both EDR and PDR data required close contact of the sensors to the skin, the skin-electrode impedance of ECG might be more easily affected by displacement due to motion artifacts. Although motion might also induce variations in the BW component PPG, the respiratory modulation of the venous blood likely dominated. Gait-related movement artifacts were also less noticeable in the PDR signal when compared to EDR signal. Second, an intrinsic advantage of the BW-based PDR signal over the AM-based EDR signal was due to their differences in sampling mechanisms and sample rate (67 Hz for PDR vs. 1~4 Hz for EDR, as ECG SR features were extracted at each heartbeat). For this reason, the PDR signal was not subject to the Nyquist sampling limit while the EDR signal was. Finally, extracting EDR signal required accurate R peak detection while the PDR signal was extracted using a simpler procedure.

4.4.3 Generalizability of the Proposed Method

MA-U-Net was shown to be generalizable to unseen modalities, subjects, and activities as evident in the results of Experiment 3. Surprisingly, the model trained on the EDR data of D_{source} generalized well to the fused EDR and PDR data of D_{target}. This was likely because EDR and PDR data were similarly distributed, regardless of subject. The results also indicated the potential to apply the MA-U-Net trained on D_{source} to a novel dataset with different modalities (e.g., impedance pneumography [118], ultra-wideband radar [130], video [131], [132], etc.), which further expanded its use cases. Note that we have also trained U-Net using SDR data of D_{source}, but we found that its performance was consistently worse across subjects. It is also critical to use a source dataset with sufficiently high quality (e.g., by using stationary activities only) to achieve better performance on the target dataset.

4.4.4 Combining the Wearable Patch and 6MWT for Assessing Cardiopulmonary Deterioration

The 6MWT could be a simple, safe, and inexpensive tool for self-administered assessment of disease severity [133]. Further, 6MWT-R was included to simulate patients who suffer from COPD. Normally, 6MWT distance was the primary data used for assessment, while we believed measuring health parameters such as RR simultaneously could strengthen the prognostic value of this tool. Unfortunately, health parameters were rarely measured due to the lack of small wearable devices to collect such data when motion artifacts were present during 6MWT [114]. This work presented hardware with a small form factor and an algorithmic system directly demonstrating accurate RR estimation during 6MWT and 6MWT-R. We believe this system can be easily combined with the existing 6MWT protocol to study cardiopulmonary deterioration in the future.

Moreover, we would also like to explore new health applications that utilize the estimated RR as a feature for disease monitoring. For example, pulse-respiration quotient (PRQ)—which is the ratio of HR to RR—was previously found to be representative of the autonomic regulatory states of the cardiopulmonary system [134]. Therefore, changes in PRQ could indicate abnormality of the autonomic nervous system. To the best of our knowledge, no wearable system has exploited this phenomenon due to the lack of suitable wearable devices for accurate HR and RR estimation robust to motion artifacts. The methods presented in this work may enable broader and more frequent usage of PRQ in less controlled settings.

4.4.5 Alternative Wearable Systems

Currently, there were several wearable devices proposed for tracking RR during walking. Raiano *et al.* [135] achieved accurate RR estimation during walking using a smart garment, which places piezoresistive sensors around the abdominal rib cage and abdomen. Impedance pneumography-based wearable systems have also shown promising results during walking in the state-of-the-art method presented by Berkebile *et al.* [118]. Chu *et al.* proposed placing small strain sensors on the abdomen and ribcage [136]. The strain applied to these sensors by respiration can be used to estimate RR even during walking. Yet, wearable systems built using piezoresistive sensors, impedance pneumography sensors, and strain sensors were generally less preferable for long-term, at-home usage since they are bulkier and/or require careful placement. Due to the sensing principles of impedance pneumography, the distance required between the sensors for accurate RR estimation may pose an unavoidable limit on the package size. Nevertheless, the diminished accuracy of the miniaturized version of these devices might be compensated by

our wearable patch, algorithmically through MA-U-Net and physically by integrating the sensing hardware.

4.4.6 Limitations and Future Work

Despite the high RR accuracy attained in a less controlled setting, there were still a few limitations. First, the subjects analyzed were healthy, and it remains uncertain whether this approach can be adapted to a clinical population. Therefore, to demonstrate the usability of this method for early prediction of life-threatening events (e.g., cardiac arrest), data from the population at risk of these events will need to be collected. Problems such as abnormal cardiac rhythm and abnormal cardiovascular signal morphology could raise new engineering challenges especially when extracting EDR signals. For these reasons, we hypothesize that PDR signals may be more robust in these cases, though further validation would be required.

Our target population—patients with cardiopulmonary complications—tend to be less active, but it is still critical to ensure accurate RR estimation when these patients are moving. Although we showed that applying MA-U-Net on EDR and PDR signals with outliers rejected could yield a low MAE, we have not shown a method that is adaptable to various activity types. Fortunately, An *et al.* [108], [109] have previously developed deep learning models for human activity recognition (HAR), capable of classifying various activities such as rest, walking, stair climbing, and running accurately using ACC signals measured from the same device. Such HAR models can be combined with MA-U-Net in a hierarchical framework: 1) Classify the activity type performed by a subject. 2) Select the activity-specific RR estimation model accordingly (e.g., *All* + SE-U-Net for stationary

activities, *EDR & PDR* + MA-U-Net for walking, and *PDR* + U-Net for running). In future work, RR estimation performance should be assessed using this hierarchical framework to understand true performance as if the models are deployed in real world.

Despite accurate RR estimation during walking, the MAEs could still be high for some subjects. The DR data of some subjects could still be jointly noisy. The sources of these noises for these subjects remain unknown. Elucidating them is the key to further improving estimation performance during walking. Besides the unknown noise sources, we should also mitigate noises with known origins. We experienced some problems with ECG during data collection and needed to reject some subjects or use ECG measured at a different body location consequently. Future studies should investigate a robust way to measure ECG using this patch when subjects are moving more.

4.5 Conclusion

A pipeline to estimate RR using data measured from a chest-worn wearable patch has been proposed. In particular, the MA fusion module in this pipeline addressed heterogeneous respiratory quality across modalities during walking. We showed that EDR, SDR, and PDR data may complement each other, depending on the activity type. During walking, EDR fused with PDR via MA-U-Net achieved the best performance when compared to the alternatives. Besides better fusion capability, and MA-U-Net could also provide easily explainable modality quality. Finally, RQI_{FFT} has also shown its effectiveness in identifying outliers. Although the algorithms were not validated in a clinical population yet, PDR signals may be generalizable even for patients with pathological conditions. Future work should also assess the performance of a hierarchical

framework that combines activity-specific RR estimation models with an HAR model. Besides these challenges, we hope to use the RR estimated from this device to investigate its prognostic power for disease prediction, detection, and tracking in the future.

In this chapter, we replaced the fusion and the denoising modules in a classical DSP pipeline for RR estimation by a DL architecture. Since the fusion module was designed using the DSP-based RQI_{FFT}, the intermediate outputs of the DL model were therefore more interpretable, more debuggable, and more logical from the signal processing standpoint. Here, we discovered a new potential direction of the replacement strategy we have taken thus far. We have noticed that: 1) including DSP-inspired functional modules can also benefit the DL model as it can improve its interpretability and 2) it may be possible to implement DSP functional modules with the building blocks of neural networks. The promising results attained in this chapter has made us wonder if the relationship between DSP and DL has begun to transition from no interactions to mutualism. In the next chapter, we will continue to explore other ways to reap the full potential of this strategy.

CHAPTER 5. USING FAST FOURIER TRASNFORM IN DEEP LEARNING TO EXTRACT FEATURES FOR HEART RATE ESTIMATION

Estimating HR from the SCG signal can provide a more unobtrusive alternative for long-term HR monitoring where the gold standard ECG signal is less favorable. DL approaches have demonstrated promise in accomplishing this and are attractive due to their flexible and data-driven nature. However, current dense layer-based DL approaches lack a carefully designed regressor to estimate the HR in the SCG input and may overfit in low-data regime. It is also uncertain how well most of these DL approaches can generalize to unseen subjects, as evaluation has primarily taken place with no separation of subjects between training and testing data. DL practitioners may fall into such pitfall if they fail to recognize the goal for using DL in this field: enabling accurate wearable devices not just for the subjects in the training dataset but also for the unseen subjects. This work can lead to accurate, real-time algorithms for estimating 60s HR from a single chest-worn tri-axial ACC only, which could be embedded in textiles.

Improper training-testing scheme can be addressed by splitting the datasets using LOSO cross validation scheme. However, it is less obvious how to design better HR regressor using DL. In this chapter, we introduced the use of FFT within a DL architecture to enhance the learning of the layers prior. This is the first time FFT is used within a DL architecture for HR estimation, and to the best of our knowledge, it may also be the first time FFT was used in this fashion for health parameter estimation. Most functional modules in a DSP pipeline including signal smoothening, ST transformation, and identification of the dominant frequency have been replaced by building blocks of neural

network in this work. The only functional module left in the DSP pipeline was the step that required the least fine-tuning: bandpass filtering. As hinted at the end of **CHAPTER 4**, other DSP-based modules (e.g., FFT and dominant frequency detector) may also be implemented in neural networks. Their DL counterparts, despite lacking trainable weights and thus were not adaptable to the data, were still helpful by imposing constraints in the model architecture. Most interestingly, the DL techniques developed for other problems such as semi-supervised learning can now be integrated with the DSP-inspired functional modules. For example, we used the domain adversarial technique previous developed by Ganin *et al.* [137] as a mean to regularize the training of SCG's feature extractors so the their deep features can be more indiscriminative to the deep features extracted from the gold standard ECG. In short, representations already enhanced by DSP-inspired block can be further improved by semi-supervised learning.

5.1 Introduction

HR has long been used for a diverse range of applications, including physical assessment [138], monitoring disease progression [9], and modeling the autonomic nervous system's response to nerve stimulation [139]. The current gold standard for deriving HR is through electrocardiogram ECG measurement [140], requiring gel electrodes that may be irritable to the skin. Measuring the mechanical vibration of the heart using ballistocardiogram (BCG) or SCG signals may offer a more comfortable alternative for long-term HR monitoring solution, as these sensors can be built into a bathroom scale [34], [141], hospital bed [142], clothes [143], or small chest-worn wearable device [35]. In particular, SCG measures the local vibrations of the chest wall due to cardiac activity [94], and may reveal underlying cardiovascular state [33].

Methods to estimate HR from BCG or SCG include signal processing [143], [144] and data-driven approaches [143], [145]–[148]. Data-driven approaches are more favorable as they can be adapted and/or specialized to different datasets or sensing modalities. Some of the more recent data-driven approaches are DL-based. One common variant [145], [147] of these methods first removes high frequency oscillations through convolutional filters intended to retain cardiac pulsation. Next, dense layers are used to regress HR. Note the recurrent cells in [147] is built on dense layers and gated, recurrent mechanism. However, we notice that these dense regressors are prone to overfitting and do not generalize well to unseen subjects in a preliminary analysis in our dataset; dense layers may contain myriad parameters, increasing the amount of data required to ensure generalizability. Unfortunately, except for [147], some of the aforementioned DL-based methods [145], [146] do not split the training and testing sets by subjects, violating the identically and independently distributed assumption [3] and misinforming the generalizability of the model. Therefore, it remains uncertain whether dense layer-based DL models are generalizable to unseen subjects' data. Assessing model performance on unseen subjects cannot be overlooked because the high inter-subject variability of SCG can hinder estimation performance more than its intrasubject variability [94].

To address this, we propose to replace dense layers with a dominant frequency regressor (DFR). In addition, we employ domain adversarial training (DAT) [137] to encourage the extraction of sensor invariant features by including a domain classifier with gradient reversal layer (GRL) that leverages ECG during training. LOSO cross validation experiments are conducted to compare different architectures. To the best of our knowledge, this is the first end-to-end DL-based model that does not utilize the

predominantly used dense regressor and the first DL approach that leverages ECG and DAT to regularize for estimating HR from SCG.

5.2 Methods

5.2.1 Dataset

In this study, we recruited 21 healthy subjects under a protocol approved by the Georgia Tech Institutional Review Board (Protocol Number H21100). Data from two subjects are excluded due to unsatisfactory ECG quality, determined by manual inspection. Only the data from the remaining 19 subjects (age range: 20-30 years old, 12 males and 7 females) are presented in this chapter. Our custom-built wearable patch—previously validated for estimation of several respiratory parameters [35], [105], [18]—was placed at the mid-sternum of the subjects. This patch measures single lead ECG, tri-axial accelerometer, and photoplethysmogram signals, but only ECG and dorso-ventral accelerometer (which will be referred to as SCG hereinafter) signals are used in this work. The details of the hardware specifications can be found in [105]. Subjects performed several activities such as sitting still, walking, running, etc., but we only assess the model performance when subject sits still in upright posture for three minutes in this work.

5.2.2 Signal Preprocessing

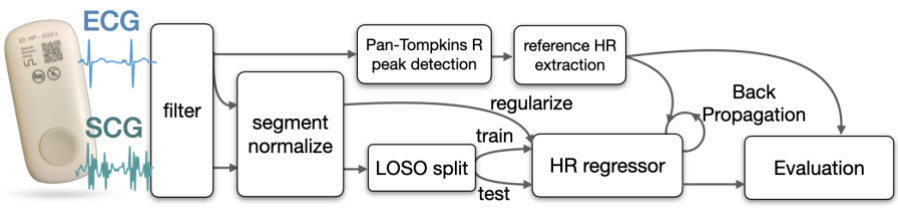
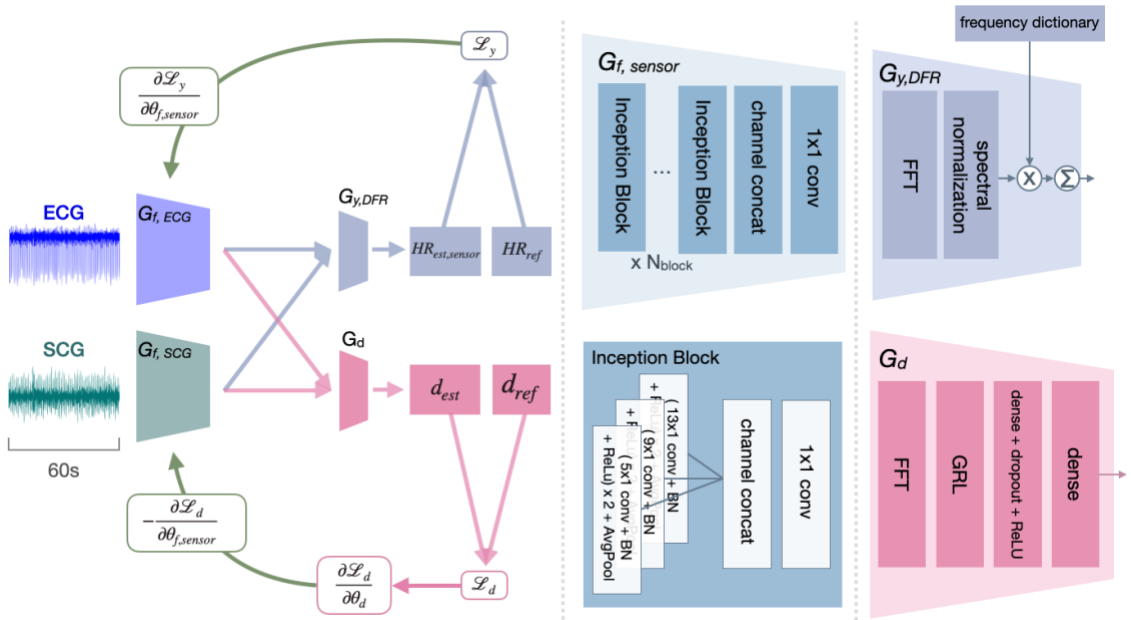


Figure 5.1. Heart rate estimation pipeline.

As illustrated in Figure 5.1, the ECG is used to provide the reference HR (HR_{ref}) for training and regularizing the DL model, and SCG is used for training and inferring HR ($HR_{est,SCG}$). ECG and SCG are first passed through 4th order Butterworth bandpass filters with passbands of 1-30Hz and 1-40Hz respectively [18]. R peaks are detected from the ECG using the Pan-Tompkins algorithm [49]. We also visually inspect all R peaks to ensure accurate R peak detection. Next, HR_{ref} is derived from the RR intervals. To prepare the data for the DL model, ECG and SCG signals are segmented into 60s windows with 90% overlap. The window of a signal is denoted as x_{sensor}, where *sensor* can be either ECG or SCG. Note that window sizes of 4 seconds [149] to 2 minutes [144] have been previously used in the literature. These SCG windows are clipped by ±5 standard deviations above and below the mean SCG to reduce motion artifacts (adapted from [150]). Next, these windows are normalized (zero mean unit variance) to account for variability in SCG amplitude. Note that no outlier rejection algorithm is used.

5.2.3 *Overview of the Deep Learning Architecture*



(a) (b) (c)

Figure 5.2. (a) Overview of the deep learning architecture. (b) Both feature extractors ($G_{f,ECG}$ and $G_{f,SCG}$) consist of N_{block} inception blocks and a 1×1 convolutional layer (channel pooling). (c) The dominant frequency regressor $G_{y,DFR}$ and the modalitiy classifier G_d with gradient reversa layer (GRL).

Figure 5.2 (a) shows the DL architecture proposed in this work. The preprocessed signals are first passed to the feature extractors, $G_{f,sensor}$. The extracted features are next passed to the dominant frequency regressor $G_{y,DFR}$ for HR estimation and to the domain classifier G_d for sensor classification.

5.2.4 *Feature Extractor based on Inception Block*

Figure 5.2 (b) shows the building block of $G_{f,sensor}$ (parameterized by $\theta_{f,sensor}$), which is adapted from the inception block in [124], [151]. We denote the number of inception blocks used as N_{block}. Each inception block starts with three parallel convolutional layers (with kernel sizes of 5, 9, 13 and a channel size of N_{ch}), following by a batch normalization

(BN) layer, and a Rectified Linear Unit (ReLU) layer. The convolutional layers, BN, and ReLU layers are repeated twice. Next, each of these three data streams are compressed using an average pooling (AvgPool) layer with the same kernel sizes as its convolutional layer upstream and a stride size of two. These data streams are concatenated along the channel dimension and reduced to N_{ch} channels through a 1×1 convolutional layer. After N_{block} inception blocks, N_{ch} channels are reduced to one channel also through a 1×1 convolutional layer. The output of $G_{f,sensor}$ is passed to $G_{y,DFR}$ and G_d.

5.2.5 Dominant Frequency Regressor

$G_{f,sensor}$ can denoise SCG and retain information only relevant to the main task: signal oscillation at the cardiac frequency. In contrast to using dense layers for HR estimation, our more interpretable approach—Dominant Frequency Regressor $G_{y,DFR}$—forces these signals to resemble sinusoidal signals by transforming the output of the convolutional layers into the spectral domain using FFT. Note that sinusoidality and more interpretable patterns can only be somewhat observed with dense regressor because dense layers allow for further computation following $G_{f,sensor}$. In Figure 5.2 (c), the extracted spectral features are masked so only data between 40 to 150 beats per minute (bpm) are preserved (110 frequency bins). This is the expected range of HR for this dataset. Signal processing approaches analogous to this procedure usually extract HR by determining the dominant frequency through the argmax function. Since argmax is non-differentiable, it is generally not used during backpropagation. To address this, we propose to first rescale the spectral features, so they sum to one. Next, we treat the spectral features as the probability densities and use them to compute the inner product between the spectral features and their corresponding frequencies (frequency dictionary in Figure 5.2 (c), in bpm). This is

equivalent to computing the expectation and is treated as $HR_{est,sensor}$. We summarize the above operations as follows,

$$HR_{est,sensor} = G_{y,DFR}\left(G_{f,sensor}\left(x_{sensor}, \theta_{f,sensor}\right)\right) \quad \textbf{(5.1)}$$

Next, we compute the regression loss $L_{y,sensor}$ using Mean Square Error between $HR_{est,sensor}$ and HR_{ref}. Note $G_{y,DFR}$ does not have trainable parameters.

5.2.6 Domain Adversarial Training

Since ECG signals are available, we also attempt to use ECG to regularize $G_{f,SCG}$. This can be accomplished by a domain classifier G_d (parameterized by θ_d) to extract domain invariant features from $G_{f,sensor}$. In the context of our work, the domain is equivalent to the sensor type of the input. As shown in Figure 5.2 (c), G_d consists of FFT transformation, a GRL, a dense layer (110 × 50), a dropout layer (p=0.5), a ReLU layer, and another dense layer (50 × 1). This G_d is adapted from [152]. G_d will classify the sensor d_{est} of the signals, and the GRL will pass the gradient multiplied by negative one to $G_{f,sensor}$. This enforces $G_{f,sensor}$ to compete with G_d in an adversarial way—the former will maximize the sensor classification loss L_d and the latter will minimize it. This could encourage $G_{f,ECG}$ and $G_{f,SCG}$ to extract common information between the sensor signals such as cardiac pulsation. We summarize the above operations as follows,

$$d_{est} = G_d\left(G_{f,sensor}\left(x_{sensor}, \theta_{f,sensor}\right), \theta_d \right) \quad (5.2)$$

Next, we compute L_d using the binary cross-entropy loss between d_{est} and the label d_{ref}.

5.2.7 Loss Functions, Model Training, and Evaluation

The total loss is expressed as follows,

$$L_{total} = L_{y,SCG} + L_{y,ECG} + \lambda L_d \quad \textbf{(5.3)}$$

Here, λ is the regularization weight for controlling the trade-off between HR estimation and DAT.

For training and testing, we use the LOSO cross validation scheme (i.e., dataset is split into 19 folds, 18 subjects in training set and one subject in testing set). During backpropagation, we update the parameters using the Adam optimizer with a learning rate of 0.001 and a weight decay of 0.01. The mini-batch size is set to 64. All models are trained for 400 epochs. We compute the MAE, RMSE, MAPE, and PCC of the optimal model on the aggregated left-out testing sets. We also report the averages and standard deviations of MAE and PCC across the subjects to compare different model architectures. Note that ECG is not used during testing and its R peaks are not used as timing references.

5.2.8 Model Architecture Comparison

Model performance can be sensitive to the regressor choice, N_{block} (depth) and N_{ch} (width) of the $G_{f,sensor}$, and λ. For the regressor, we implement a dense regressor $G_{y,dense}$ built on dense layers. $G_{y,dense}$ consists of a BN layer, a ReLU layer, a dense layer (dimension of $G_{f,sensor}$ output × 50), another BN layer, another ReLU layer, and another dense layer (50 × 2). To make the comparison fair, several model depths for $G_{y,dense}$ are attempted. In total, ten model architectures are studied.

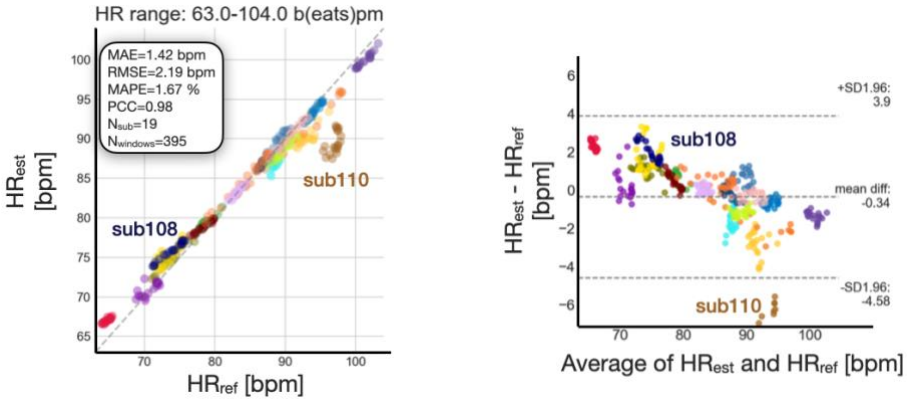
1-5) **Impact of λ:** $G_{y,DFR}$, N_{block}=3, N_{ch}=8, and λ=[0, 1, 5, 10, 50].

6) **Impact of N_{block}:** $G_{y,DFR}$, N_{block}=2, N_{ch}=8, and λ=0

7) **Impact of N_{ch}:** $G_{y,DFR}$, N_{block}=3, N_{ch}=4, and λ=0

8-10) **Impact of $G_{y,dense}$ with various N_{block}:** $G_{y,dense}$, N_{block}=[3, 6, 8], N_{ch}=8, and λ=0

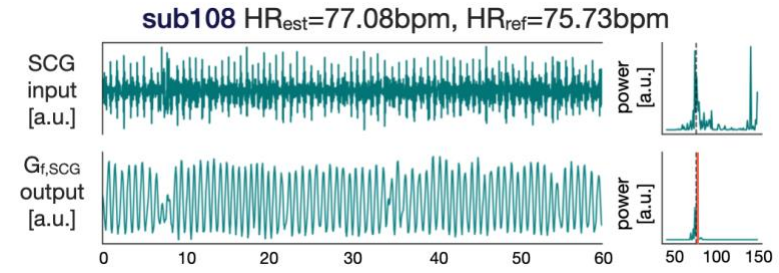
5.3 Results and Discussions



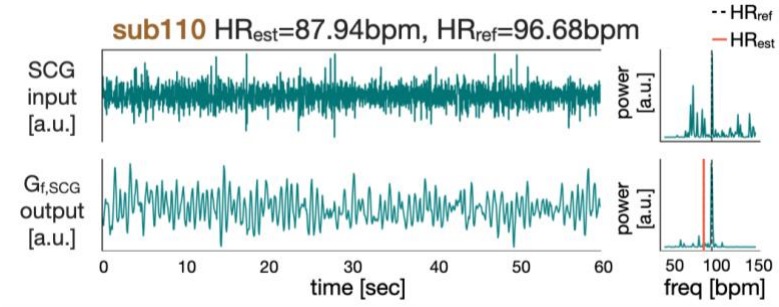
(a) (b)

Figure 5.3. (a) Regression analysis of HR estimated from SCG ($HR_{est,SCG}$) compared to the reference HR derived from ECG (HR_{ref}). (b) Bland-Altman analysis of $HR_{est,SCG}$ compared to HR_{ref}. Note the colors in the plots are used to separate the subjects.

The model architecture with the lowest average MAE (model architecture 4) has the following hyperparameters: $G_{y,DFR}$, N_{block}=**3**, N_{ch}=**8**, and λ=**10**. The inference speed of this model on a CPU (Intel® Core™ i7-7820X @ 3.60 GHz) is approximately 5.7ms ± 153µs (tested on a 60s SCG window, experimented 1000 times). As depicted in Figure 5.3 (a), this model achieves an MAE of 1.42 bpm, an RMSE of 2.19 bpm, an MAPE of 1.67%, a PCC of 0.98 on the aggregated left-out testing sets. As shown in Figure 5.3 (b), the 95% limits of agreement are -4.58 to 3.9 bpm. We highlight two representative subjects: subject 108 (good quality) and subject 110 (poor quality) and analyze them in the subsequent section.



(a)



(b)

Figure 5.4. $G_{f,SCG}$ (model architecture 4) output of an SCG window. (a) Clean output (AE = 1.35 bpm) (b) Noisy output (AE = 8.74 bpm). Left side: temporal domain. Right side: spectral domain.

In its simplest form, cardiac pulsation can be represented by signals that resemble sinusoidal waves. In Figure 5.4 (a), we show that $G_{f,SCG}$ is capable of removing morphological characteristics of SCG for this test subject. The resulting output is a clean sinusoidal wave oscillating at the cardiac frequency. However, in Figure 5.4 (b), noises contaminate SCG more severely, and $G_{f,SCG}$ is not able to reduce these noises fully for another test subject.

Table 5.1 Estimation performance of different architectures.

Gy	Gf	λ	MAE [bpm]	PCC	Params
$G_{y,\,DFR}$ (proposed)	N_{block} =3 N_{ch} =8	0	1.80 (1.61)	0.73 (0.45)	9.9k
		1	1.74 (2.23)	0.77 (0.34)	25.4k
		5	1.58 (1.71)	0.80 (0.29)	25.4k
		10	**1.42 (1.66)**	**0.89 (0.15)**	25.4k
		50	1.87 (1.96)	0.76 (0.43)	25.4k
	N_{block} =2 N_{ch} =8	0	2.65 (2.99)	0.66 (0.44)	6.1k
	N_{block} =3 N_{ch} =4	0	2.02 (2.28)	0.72 (0.51)	2.6k
$G_{y,dense}$	N_{block} =3 N_{ch} =8	0	14.41 (6.64)	-0.04 (0.44)	49.1k
	N_{block} =6 N_{ch} =8	0	4.60 (2.59)	0.47 (0.36)	26.4k
	N_{block} =8 N_{ch} =8	0	7.16 (6.31)	0.16 (0.48)	30.3k

The results in Table 5.1 shows the performance of the model architectures described in Section 5.2.8 in the same order. The carefully designed $G_{y,DFR}$ is clearly better than $G_{y,dense}$. This is likely because $G_{y,DFR}$ makes realistic assumption on the structure of the data, which dense layers may not be able to learn in low-data regime. We also notice that higher N_{block} and N_{ch} can help improve estimation performance. This implies that the model may be underfitting if N_{block} and N_{ch} are too small.

The regularizer weight λ needs to be selected carefully since some λ can lead to undesired negative transfer (e.g., λ = 50). There exists a trade-off between extracting features that are specific to the target task (i.e., HR estimation) and those that are sensor invariant. Low λ will reduce the alignment between both signals. High λ may lead to extraction of information that is common between the modalities but not necessarily driven by cardiac pulsation (e.g., respiratory modulation).

The proposed model has a few limitations. First, the model estimate HR from a 60s SCG window and therefore does not provide beat-by-beat HR. Sudden changes in HR (e.g., during exercise recovery) may not be fully captured. Further, the model is only assessed on seated subjects who do not show high intrasubject variability in HR during recording. Since SCG measured at different sensor locations may have different data distribution for certain subjects [153], the convolutional layers may require further fine-tuning. On the other hand, this model shows promising results and may be applicable to other HR estimation problems that use different sensing modalities such as RGB video [154] or impedance sensors [155]. Regularizing using ECG is also attractive as validating HR estimation already requires simultaneous ECG measurements to compute the ground truth HR in most cases. Further, the extracted features may be used to improve downstream tasks such as oxygen uptake estimation from SCG [35]. Finally, the model can also be adapted for DL-based respiratory rate estimation algorithm such as [18], though the $G_{f,sensor}$ may need to be deeper to capture respiratory modulation.

5.4 Conclusion

In this work, we proposed a DL model with a novel dominant frequency regressor $G_{y,DFR}$ and regularize it using DAT for HR estimation. This model is superior to the predominantly used dense regressor and selecting the right λ can further improve the estimation performance. One future direction of this work includes adapting the model for different sensing modalities and regression tasks. Further, we also need to assess the model performance when HR changes more abruptly.

In **CHAPTER 4** and **CHAPTER 5**, we have presented two examples that replaced multiple functional modules in a DSP pipeline. We also discovered freshly uncovered advantages that were not obvious before—the mutualism relationship between DSP and DL, the feasibility to implement DSP methods using neural networks (and how doing so is helpful for the DL model), and the compatibility of the DL regularization techniques with DL functional modules heavily inspired by DSP. The progress from **CHAPTER 2** to **CHAPTER 5** has prepared us well for addressing more complicated problems that may suffer from our poor understanding of the relationship between the physiological signals and the target health parameter, the difficulty to provide intuitive meaning of the features extracted by DL models, the lack of consideration toward the heterogenous signal quality discussed in **CHAPTER 4**, etc. In the next chapter, we will study one such problem while further dissolve the boundary between DSP and DL, striking a balance among accuracy, interpretability, and logicality.

CHAPTER 6. MERGING PHYSIOLOGY AND DIGITAL SIGNAL PROCESSING INTO A DEEP LEARNING ARCHITECTURE FOR VO_2 ESTIMATION

Monitoring VO_2 using indirect calorimetry from critically ill patients can help reveal underlying pathological conditions and inform treatment efficacy [156]. However, the use of indirect calorimetry is not suitable since it is inconvenient to use. Most wearable solutions developed to estimate VO_2 attempted to relate the activity level and HR to VO_2, but they fail to consider the cases where variations of VO_2 may not depend on the movement of the body. The chest-worn wearable patch used throughout this dissertation is unobtrusive to wear while it could also capture information related to VO_2 that did not depend on the activity level.

Besides addressing the clinical problem, we also addressed **Aim 3** by merging DSP and physiological knowledge into a DL model for improving VO_2 accuracy in this chapter. A DL architecture was proposed to estimate VO_2 from filtered SCG, ECG, and PPG, which was composed of layers of convolutional filters, the FFT-based functional module presented in **CHAPTER 5**, a channel attention module and a spectral attention module inspired by common DSP functional modules as well as the modality-attentive fusion presented in **CHAPTER 4**, and a dominant frequency regressor presented in **CHAPTER 5**. Most importantly, as motivated by the domain adversarial regularizer used in **CHAPTER 5** that exploited the additional **signals** available in the dataset, we also exploited the additional **labels** in this dataset through multi-task learning (MTL) regularization. Through MTL, we found that domain knowledge in physiology naturally informed the optimal auxiliary tasks. Because of the way we formulated the MTL

framework, the auxiliary tasks, which acted as regularizers, also enforced extraction of intuitive features.

6.1 Introduction

Monitoring VO_2 using indirect calorimetry from critically ill patients can help determine the degree of hypermetabolism, which may reveal underlying pathological conditions and inform treatment efficacy [156]. However, the use of indirect calorimetry is not suitable for tracking VO_2 frequently since it requires a cumbersome calibration procedure and is obtrusive to patients since face mask needs to be worn. Hence, wearable form factors that are noninvasive and unobtrusive would be more feasible to address the convenience problem. Most wearable solutions proposed primarily rely on relating the activity level (through ACC sensors) and HR to VO_2. Nevertheless, one of the assumption these approaches depend on is usually not true for the target population: the variations of VO_2 are due to the movement of the body. Indeed, VO_2 can change due to immune and inflammatory responses even when patients are not moving at all. To address this, we explored cardiorespiratory information that can be extracted from the SCG, ECG, and PPG measured by the chest-worn wearable patch for estimating VO_2. This method will capture the information that are related to VO_2 more directly and will not depend on the activity level.

Most wearables that can monitor cardiopulmonary functions have their own limitations, rendering them unsuitable for VO_2 tracking. First, the most convenient form factor, the smart watch, mostly used HR information for stationary VO_2 estimation, but inferring VO_2 from HR alone can be less accurate than the ventilation approach [157]. On the contrary, the ventilation-based wearables [158]–[160] (which may sometime be accompanied with cardiac information [161]) usually have larger footprint (e.g., smart shirt). To exceed the adoption threshold, wearables need to be convenient and accurate,

which may be both addressed by the chest-worn wearable patch. Specifically, we believe this patch can improve the accuracy compared to other wearables because of the rich cardiac and respiratory information it can offer. Note that we will reference literatures related to VO_2 and EE interchangeably since in this work, the EE derived using VO_2 and VCO_2 through the Weir's equation [162] has a near linear relationship.

Estimating VO_2 from the cardiac and respiratory information in wearable physiological signals would rely on the Fick's law and the sensing principles of indirect calorimetry respectively. The CO, which is the multiplication of HR and SV, have shown a strong correlation yet sublinear relationship (due to the dynamic oxygen delivery efficiency to tissue) with VO_2. On the other hand, the ground truth device used in this work (despite exhibit a small error when compared to Douglas bag method), COSMED K5, measured RR, TV, O_2 fractional concentration in inspired and expired air, etc. simultaneously to compute VO_2. Further, MV (ventilation), which is the multiplication of RR and TV, have also shown a strong correlation and linear relationship with VO_2 [5], [163]. Encouraged by these findings as well as the salient SCG features for estimating TV [19] and for estimating SV [15], [16], we studied methods to extract the features that carry information about HR, RR, SV, and TV from the wearable patch.

In this work, we proposed a DL architecture to extract cardiac and respiratory information for VO_2 estimation. This architecture was designed based on popular DSP techniques for extracting relevant features and domain knowledge in physiology. The contributions include the following: 1) proposed an DL architecture to extract and decouple interpretable cardiac and respiratory information in SCG through FFT and MTL for VO_2 estimation, 2) proposed interpretable channel and spectral attention to better fuse and denoise the extracted features, 3) demonstrated the complementariness of the SCG-derived cardiac and the respiratory features toward VO_2 estimation for the first time.

6.2 Methods

6.2.1 Dataset

We used the same dataset protocol described in **CHAPTER 4** and **CHAPTER 5**. For this study, we only used the rest and exercise recovery (after 6MWT, 6MWT-R, stair climbing, self-paced walking, and running) periods. The ground truth device, COSMED K5 system, reported limits of agreement of -1.58±100.395 mL/min for their VO_2 measurements [88]. We calibrated the K5 system by following the procedure suggested by COMSED prior to collecting data from each subject.

6.2.2 Signal Pre-processing

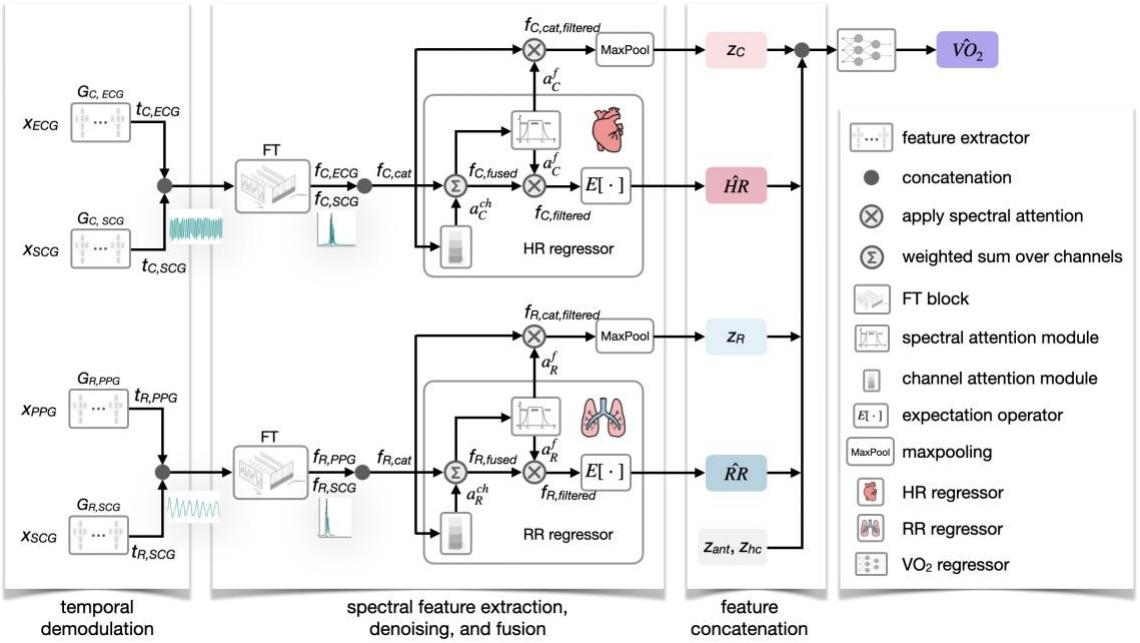
To remove out-of-band noises, we first used a fourth order Butterworth filter with a bandwidth of [0.08, 1] Hz to filter ECG signal. Tri-axial SCG signals were extracted from ACC signals using a fourth order Butterworth filter with a bandwidth of [0.08, 1] Hz. Respiratory PPG signals were filtered using a fourth order Butterworth filter with a bandwidth of [0.08, 1] Hz (or equivalent, [5, 60] brpm) to isolate the BW component of PPG. In this work, we will refer to the channel as the unique measurements within the same modality (e.g., different axis of the tri-axial SCG). The channels used include the one-lead ECG (one channel), tri-axial SCG (three channels), and red and infrared PPG on left and right sides of the patch (four channels). These signals were segmented with a 60s window and a 90% overlap. The ground truth VO_2 labels from COSMED K5 system were smoothened with a 5-point moving average filter to reduce noise and next divided by subject's weight (in kilogram [kg]) for normalization. The VO_2 were resampled to 250Hz,

and the average for each 60s window frame was computed and aligned with the wearable physiological signals.

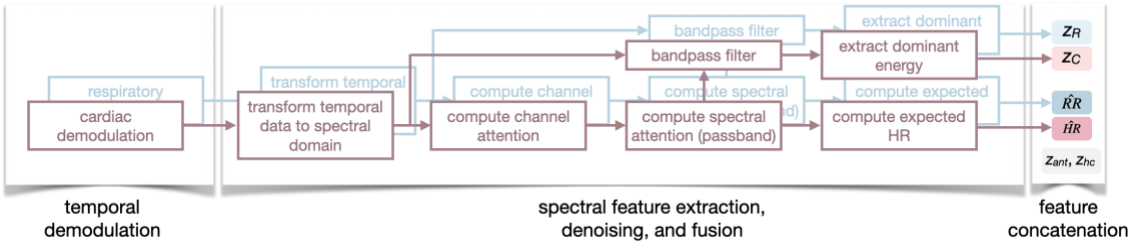
6.2.3 End-to-End Deep Learning Model for VO_2 Estimation

Here, we introduced CardioRespNet—an end-to-end deep learning model to estimate VO_2 from SCG, ECG, and PPG. The design of this model was inspired by DSP methods for RR estimation, HR estimation, VO_2 estimation, physiological signal processing and feature engineering, and our domain knowledge about the data. Further, the model was designed with improved interpretability since doing so can encourage easier troubleshooting or utility in practice [26]. After all, providing the confidence on the estimation quality is of great importance to improve public adoption [120].

6.2.3.1 Overview of CardioRespNet and Its Functional Blocks



(a)

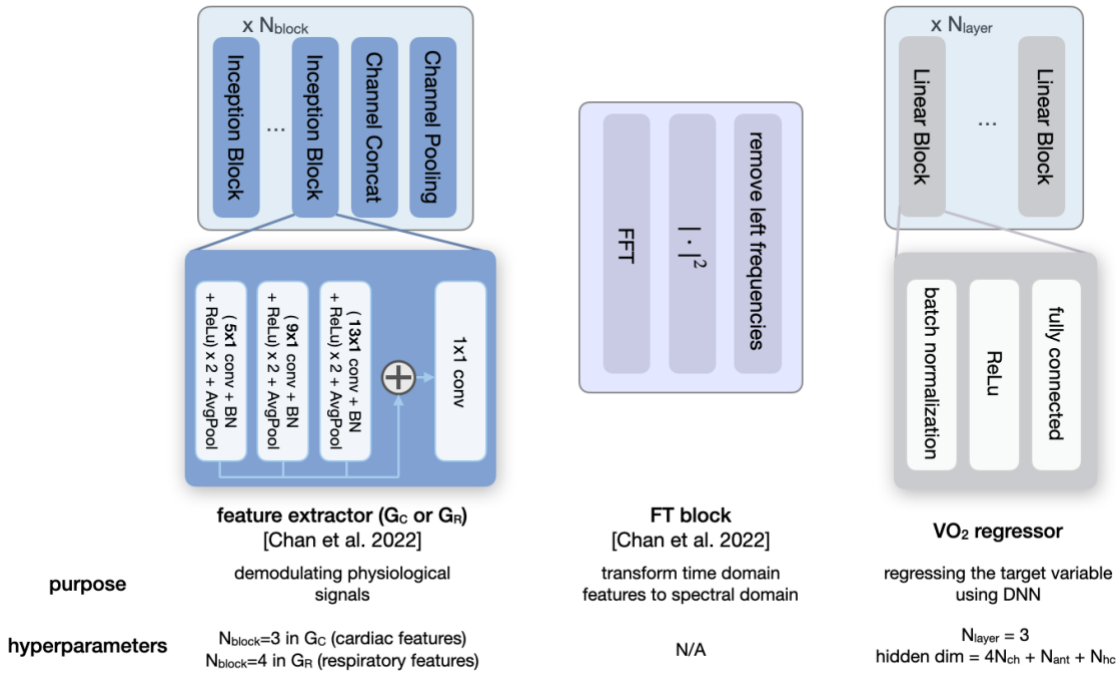


(b)

Figure 6.1. Overview of the end-to-end CardioRespNet to estimate VO_2. In this model, we applied multi-task learning (MTL), channel attention (CA) and spectral attention (SA), and HR and RR dominant frequency regressors to regularize VO_2 estimation, lending knowledge from cardiopulmonary physiology and digital signal processing. (a) Model architecture diagram. (b) Functional block diagram.

The overview of the model was presented in Figure 6.1 (a). The input data were ECG (denoted as x_{ECG}), SCG (denoted as x_{SCG}), and PPG (denoted as x_{PPG}) signals in temporal domain. The first step was temporal feature extraction, in which we applied four dedicated feature extractors to demodulate the signals. Since the features were extracted using convolutional layers only, they were considered to carry similar characteristics of the time domain signals. Therefore, we will refer to the outputs of the feature extractors as the deep temporal features hereinafter. Next, through the FT block, the deep temporal features were transformed to the spectral domain via FFT. We will refer to the outputs of the FT block as the deep spectral features. The deep spectral features were subsequently fed into the auxiliary regressors to estimate HR (middle top in Figure 6.1 (a)) and RR (middle bottom Figure 6.1 (a)). Both HR and RR regressors computed the channel attention to adaptively fuse the multi-modal data and the spectral attention to better attenuate non-cardiac or non-respiratory noises. Besides helping the HR and the RR regressors, the spectral attentions were also used to help the main task: VO_2 regression. The spectral-enhanced deep cardiac and respiratory features were finally fused with estimated HR (HR_{est}), estimated RR (RR_{est}) handcrafted SCG features (z_{hc}), anthropometric features (z_{ant}) to estimate the target variable— $VO_{2,est}$). The functional block diagram Figure 6.1 (b) was also provided to summarize the functional meanings of each block.

6.2.3.2 Multi-Modal, Multi-Task Feature Extractors



(a) (b) (c)

Figure 6.2. Functional blocks of CardioRespNet. (a) Feature extractors for demodulating the physiological signals through inception blocks and channel pooling. (b) Fourier Transform (FT) block for transforming deep temporal features to deep spectral features. (c) VO_2 regressor for estimating VO_2 using fully connected layers, batch normalization layers, and ReLU layers. DNN: dense neural network.

The first functional blocks in CardioRespNet were the feature extractors. As illustrated Figure 6.1 (a), four dedicated feature extractors were first applied: $G_{C,ECG}$ (which extracted cardiac-related temporal features $t_{C,ECG}$ from x_{ECG}), $G_{C,SCG}$ (which extracted cardiac-related temporal features $t_{C,SCG}$ from x_{SCG}), $G_{R,PPG}$ (which extracted respiratory-related temporal features $t_{R,PPG}$ from x_{PPG}), and $G_{R,SCG}$ (which extracted respiratory-related temporal features $t_{R,SCG}$ from x_{SCG}). These feature extractors shared very similar architectures, but the parameters were not shared. As depicted in Figure 6.2 (a), the building blocks of the features extractors were the inception blocks [151], which were adopted here to extract features using various kernel sizes. Similar to the designed

described in [106], the kernel sizes were 5, 9, and 13, and the stride was 2, which reduced the dimension of the signals by half if one inception block was added (or equivalently, reduced the sampling rate by half). In a preliminary analysis, we noticed that by increasing the depths of the feature extractors (i.e., N_{block}), we can extract features that reside in lower frequency range. In fact, selecting $N_{block} = 3$ was most ideal for $G_{C,ECG}$ and $G_{C,SCG}$ (sampling rate was reduced to 1/8), and selecting $N_{block} = 4$ was most ideal for $G_{R,PPG}$ and $G_{R,PPG}$ (sampling rate was reduced to 1/16). Inception blocks were selected because the various kernel sizes can capture cardiac and respiratory temporal patterns with varying lengths. The features were computed as follows:

$$t_{C,ECG} = G_{C,ECG}(x_{ECG}) \tag{6.1}$$

$$t_{C,SCG} = G_{C,SCG}(x_{SCG}) \tag{6.2}$$

$$t_{R,PPG} = G_{R,PPG}(x_{PPG}) \tag{6.3}$$

$$t_{R,SCG} = G_{R,SCG}(x_{SCG}) \tag{6.4}$$

where $x_{ECG} \in \mathbb{R}^{N_T \times 1}, x_{SCG} \in \mathbb{R}^{N_T \times 3}, x_{PPG} \in \mathbb{R}^{N_T \times 4}$, and $t_{C,SCG}, t_{C,PPG} \in \mathbb{R}^{N_{ch} \times \frac{N_T}{8}}, t_{R,SCG}, t_{R,PPG} \in \mathbb{R}^{N_{ch} \times \frac{N_T}{16}}$, N_{ch} is the number of channels, and N_T is the number of datapoints in a 60s recording. Upon inspecting the outputs at each depth of the feature extractors, we also noticed that the higher harmonics were gradually attenuated. This is understandable since the downstream tasks, (i.e., HR and RR estimation), only required information necessary to estimate the fundamental frequency. The feature extractors thus were forced to learn kernels for such goal.

We can consider the extraction of these features as a signal demodulation procedure. By applying lowpass filters, we can remove the harmonics in SCG signals and recovered the envelope that captured the fundamental frequency (equivalent to subject's

HR). If we proceed to further apply lowpass filter on the signals, we can recover the underlying respiratory signal that modulate the amplitude of SCG. Note that these processing steps may not remove the frequency modulation of the signal resulted from respiratory sinus arrhythmia.

6.2.4 Fourier Transform (FT) Block

The FT block was used to transform deep temporal features to deep spectral features. As depicted in Figure 6.2 (b), FFT was first used to transform deep temporal features to spectral domain. The magnitude of the output was next computed from the complex exponentials. Since the negative frequencies were conjugate symmetry of their positive counterparts, we only preserved the positive frequencies. Cardiac-related spectral features from x_{ECG}, cardiac-related spectral features from x_{SCG}, respiratory-related spectral features from x_{PPG}, respiratory-related spectral features from x_{SCG} were denoted as $f_{C,ECG}$, $f_{C,SCG}$, $f_{R,PPG}$, and $f_{R,SCG}$ respectively, and they were computed as follows:

$$f_{C,ECG} = FT(t_{C,ECG}) \tag{6.5}$$

$$f_{C,SCG} = FT(t_{C,SCG}) \tag{6.6}$$

$$f_{R,PPG} = FT(t_{R,PPG}) \tag{6.7}$$

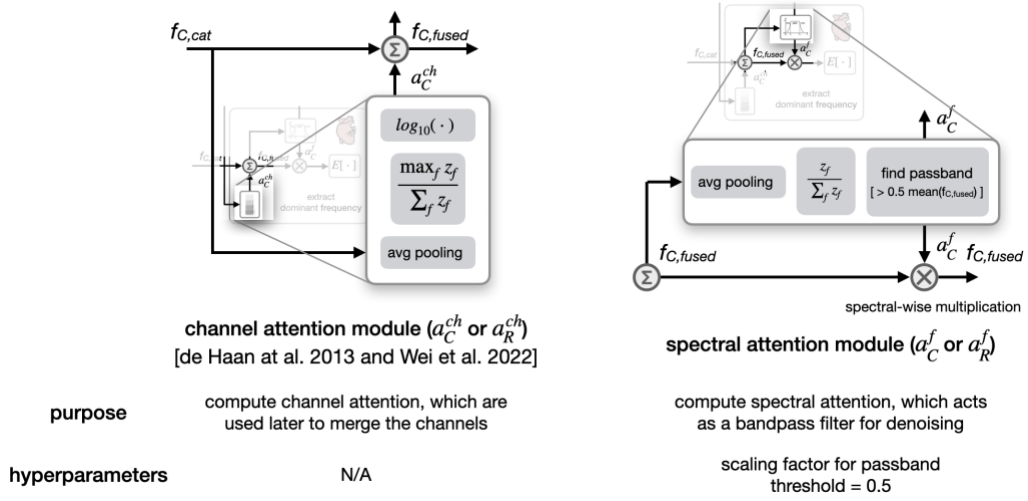
$$f_{R,SCG} = FT(t_{R,SCG}) \tag{6.8}$$

where $f_{C,SCG}, f_{C,PPG} \in \mathbb{R}^{N_{ch} \times \frac{N_T}{16}}$, $f_{R,SCG}, f_{R,PPG} \in \mathbb{R}^{N_{ch} \times \frac{N_T}{32}}$. Note that the dimensions were further reduced by half. The deep spectral features were next concatenated along the channel dimension into concatenated cardiac spectral features ($f_{C,cat}$) and concatenated respiratory spectral features ($f_{R,cat}$) as shown below,

$$f_{C,cat} = concat^{c}\left(f_{C,ECG}, f_{C,SCG}\right) \tag{6.9}$$

$$f_{R,cat} = concat^{c}\left(f_{R,PPG}, f_{R,SCG}\right) \tag{6.10}$$

where $f_{C,cat} \in \mathbb{R}^{2N_{ch} \times \frac{N_T}{16}}$, $f_{R,cat} \in \mathbb{R}^{2N_{ch} \times \frac{N_T}{32}}$. These concatenated features will be used for deriving channel attention and estimating HR, RR, and ultimately VO_2.



(a) (b)

Figure 6.3. Attention mechanisms in CardioRespNet. (a) Channel attention module for deep spectral features. (b) Spectral attention module for deep spectral features, which define the passband of the deep spectral features.

6.2.5 Channel Attention Mechanism

The purpose of the channel attention (see Figure 6.3 (a)) was to adaptively weigh the deep spectral features from each channel. The design of the channel attention module was inspired by a signal-to-noise ratio (SNR) method proposed to estimate the quality of remote PPG (rPPG) extracted from RGB video [112]. This SNR method was used later in another work to fuse rPPG signals extracted from various regions of interest [119]. As shown in Figure 6.2 (d), the deep spectral features were first smoothened using average pooling layer with a kernel size of five. Next, we divided the maximum spectral energy by

the sum of all spectral energies, which gave us the relative strength of the dominant component. By computing its natural logarithm, the SNR of a channel can be calculated. Finally, the channel attention a_R^{ch} can be computed using the softmax layer along the channel dimension as follows,

$$a_C^{ch} = softmax(SNR(f_{C,cat})) \tag{6.11}$$

$$a_R^{ch} = softmax(SNR(f_{R,cat})) \tag{6.12}$$

where $a_C^{ch}, a_R^{ch} \in \mathbb{R}^{2N_{ch}\times 1}$. The channel attentions were next used to fuse deep cardiac and respiratory spectral features (namely, $f_{C,fused}$ and $f_{R,fused}$) by weighted averaging as follows,

$$f_{C,fused}(i) = (a_C^c)^T \frac{f_{C,cat}(i)}{\sum_j f_{C,cat}(j)} \tag{6.13}$$

$$f_{R,fused}(i) = (a_R^c)^T \frac{f_{R,cat}(i)}{\sum_j f_{R,cat}(j)} \tag{6.14}$$

where $f_{C,fused} \in \mathbb{R}^{\frac{N_T}{16}}, f_{R,fused} \in \mathbb{R}^{\frac{N_T}{32}}$, and i and j denoted the index of a frequency. $f_{C,fused}$ and $f_{R,fused}$ will only be used to estimate HR and RR.

6.2.6 Spectral Attention Mechanism

The purpose of the spectral attention was for enhancing signal of interest through adaptive bandpass filtering. As shown in see Figure 6.3 (b), an average pooling layer with a kernel size of five was first applied to smoothen a deep spectral feature. Next, the deep spectral feature was divided by its sum along the spectral dimension. To determine the passband (frequencies that will be preserved) and the stopband (frequencies that will be rejected), the threshold was determined as half of the average of the spectral features. The

scaling factor, 0.5, was determined heuristically. The spectral attention for $f_{C,fused}$ and $f_{R,fused}$ (i.e., a_C^f and a_R^f) can finally be computed as follows,

$$a_C^f = [f_{C,fused} > 0.5mean(f_{C,fused})] \tag{6.15}$$

$$a_R^f = [f_{R,fused} > 0.5mean(f_{R,fused})] \tag{6.16}$$

where $a_C^f \in \{0,1\}^{\frac{N_T}{16}\times 1}$, $a_R^f \in \{0,1\}^{\frac{N_T}{32}\times 1}$, and the $[\cdot]$ operator is the Heaviside function. a_C^f and a_R^f filtered $f_{C,fused}$, $f_{C,cat}$ $f_{R,fused}$ and $f_{R,cat}$ respectively through element-wise multiplication along the spectral dimension and acted as binary masks:

$$f_{C,fused,filtered} = f_{C,fused} \odot a_C^f \tag{6.17}$$

$$f_{C,cat,filtered} = f_{C,cat} \odot (\mathbf{1}_{2N_{ch}}(a_C^f)^T) \tag{6.18}$$

$$f_{R,fused,filtered} = f_{R,fused} \odot a_R^f \tag{6.19}$$

$$f_{R,cat,filtered} = f_{R,cat} \odot (\mathbf{1}_{2N_{ch}}(a_R^f)^T) \tag{6.20}$$

where $\mathbf{1}_{2N_{ch}}$ was a 1-vector of $2N_{ch}$ rows and 1 column, and the $\odot$ operator was element-wise matrix multiplication. Applying spectral attention was expected to improve HR, RR, and VO_2 estimation tasks.

6.2.7 Auxiliary Tasks: HR and RR estimation

We applied the DFR proposed in [106] to estimate the dominant frequency (i.e., HR and RR) in $f_{C,fused,filtered}$ and $f_{R,fused,filtered}$. The first step in DFR was to reject the frequencies outside of the HR range (i.e., [40, 190] bpm) or the RR range (i.e., [5, 60] brpm) respectively. These ranges were selected based on the distributions of HR and RR

in this dataset, and they were also physiological plausible ranges during exercise recovery. Next, $f_{C,fused,filtered}$ and $f_{R,fused,filtered}$ (which may be considered as probability density function as they summed to one) were subsequently used with their corresponding frequencies to compute the expectation through dot product. This is considered a "softer" approach to estimate the dominant frequency, which in contrast to the "winner-take-all" operation through argmax, is a more differentiable generalization similar to softmax [164].

Learning these two auxiliary tasks with the main task has two benefits: 1) they helped regulate the feature extractors (i.e., $G_{C,ECG}$, $G_{C,SCG}$, $G_{R,PPG}$, and $G_{R,SCG}$), 2) the estimated HR and RR can be directly used as features for the main task as evident in a prior study [165] and 3) the by-product, a_C^f and a_R^f, can be used to enhance the spectral features of the main task. Here, we pointed out that the auxiliary regressors utilized the deep spectral features by identifying their dominant frequency. It is therefore natural to consider the use of their amplitude information for other purposes.

6.2.8 Deep Cardiac and Respiratory Features

In the previous section, the expectation of the deep spectral features was used as the dominant frequency while we believe their dominant energies may have other utilizations; for example, they can be representative of the cardiac or respiratory strengths. To further remove redundant information in the filtered spectral features, we applied the maxpooling layer on $f_{C,cat}$ and $f_{R,cat}$, which selected only the dominant spectral energies. Note that we used the concatenated features instead of the fused features because we hope to retain the expressive representations generated by the feature extractors. These deep cardiac and respiratory features were denoted as z_C and z_R, where $z_C \in \mathbb{R}^{2N_{ch}}$, $z_R \in \mathbb{R}^{2N_{ch}}$. These features can be considered as the amplitude of the fundamental components of the smoothened, filtered input signals in spectral domain, which may be related to SV [15], [16] or TV [19].

6.2.9 Main Task: VO_2 estimation

To estimate VO_2, we first concatenated z_C, z_R, HR_{est}, RR_{est}, and z_{hc} (handcrafted features, including demographic data). The concatenated features were next fed into a dense neural network (DNN) to compute VO_2. As illustrated in Figure 6.2 (c), this neural network was composed of three linear blocks, each was consisting of a batch normalization layer, a ReLU layer, and a fully connected layer. The hidden dimension is equivalent the dimension of the input to VO_2 regressor (i.e., $4N_{ch} + N_{ant} + N_{hc}$). Since XGBoost may achieve superior performance than neural network for regression task, we also studied whether replacing DNN with XGboost model improved performance. The hyperparameters were designed based on [35]. MTL was achieved by computing the Mean Squared Error (MSE) losses for HR, RR, and VO_2 estimation the gradients simultaneously, weighted by λ, λ, and $1 - 2\lambda$ respectively. Note that λ can only be between 0 to 0.5.

6.2.10 Training and Testing

In this work, we also used LOSO cross validation to assess generalizability of the models trained. To account for the randomness of initialization of the ML models, we repeated each experiment five times with differing random seed numbers.

6.3 Results

6.3.1 Accuracy of VO_2, HR, and RR Estimation

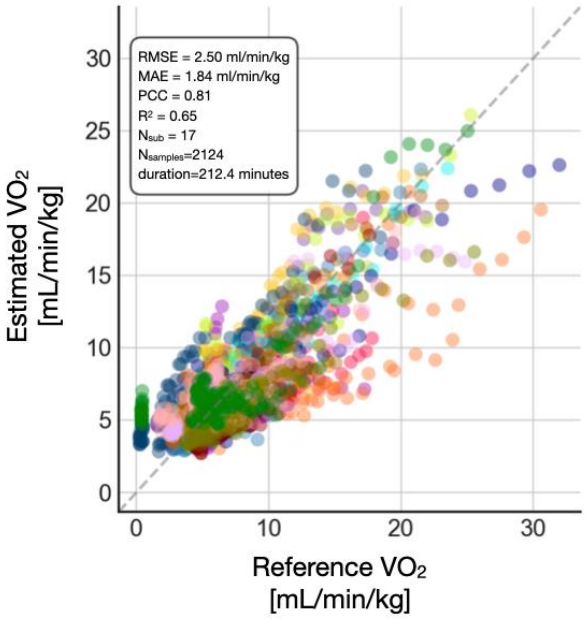
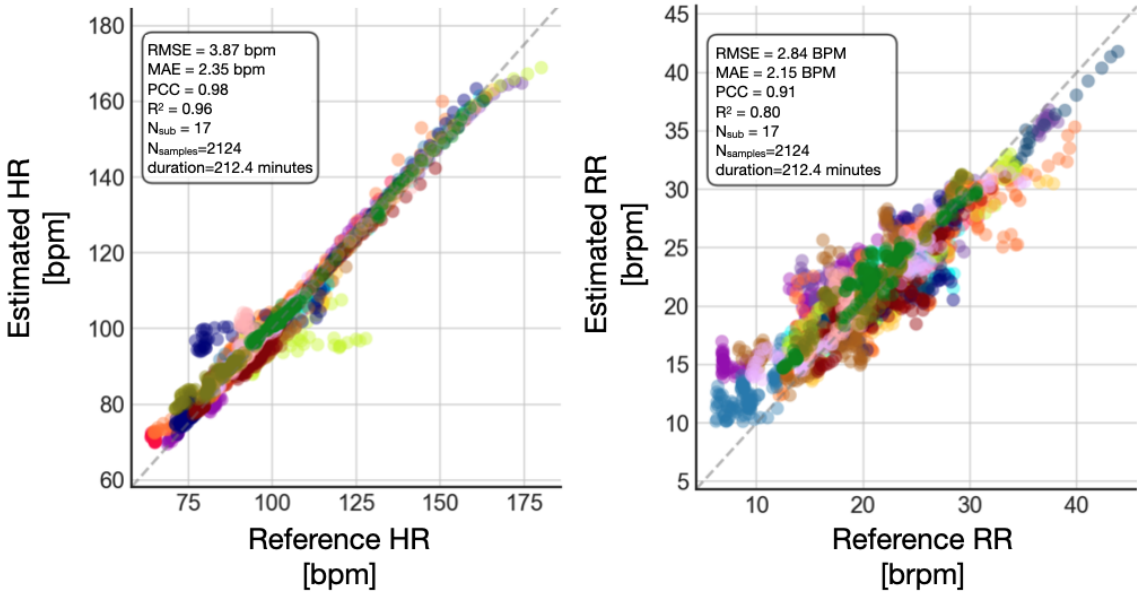


Figure 6.4. Correlation analysis for VO_2 estimation.



(a) (b)

Figure 6.5. Correlation analysis. (a) HR estimation. (b) RR estimation.

As shown in Figure 6.4 (a), we achieved a low VO_2 estimation error when compared to the ground truth using CardioRespNet. The RMSE was 2.50 ml/min/kg, the MAE was 1.84 ml/min/kg, the PCC was 0.81, the R^2 was 0.65. Further, we also achieved accurate accuracy for the auxiliary tasks. The HR and RR estimation results were shown in Figure 6.4 (b), with a R^2 of 0.96 and a R^2 of 0.80 respectively. Although the study protocol, sensing system, and algorithm were not directly comparable, we still achieved a performance comparable to the state-of-the-art method that used a activity-specific model during recovery presented in [161], which used a sensing system with a more cumbersome form factor.

6.3.2 Ablation Study

Table 6.1 Ablation study to study the impact of each function block. The number of parameters and estimation performance (RMSE) on VO_2, HR and RR estimation were presented. The FT block, MTL, CA, and SA were incrementally added. Note that λ=0.3 for this analysis.

FT block	MTL	CA	SA	Number of parameters	Estimation Performance		
					VO_2	HR	RR
×	×	×	×	3,758,167	3.91±0.31	-	-
✓	×	×	×	68,029	3.30±0.16	-	-
×	✓	×	×	3,758,167	4.01±0.49	8.12±0.59	4.71±0.93
✓	✓	×	×	68,029	2.86±0.23	**3.60±0.22**	3.77±0.23
✓	✓	✓	×	68,029	2.91±0.28	4.05±0.31	**3.31±0.31**
✓	✓	✓	✓	68,029	**2.78±0.16**	4.15±0.47	3.62±0.29

As shown in Table 6.1, FT block has helped improved estimation performance and also drastically reduced the model capacity (1/6). However, MTL, when used alone, actually hurt the estimation performance. When both were applied, an improvement of 26.85% was achieved, suggesting FT block can bring out the regularization capability of

MTL. The addition of CA and SA together also improved VO_2 accuracy marginally. Further, the addition of FT block also improved HR and RR accuracy.

6.3.3 Interpretable Features Were Enabled by FT Block and Enhanced by MTL

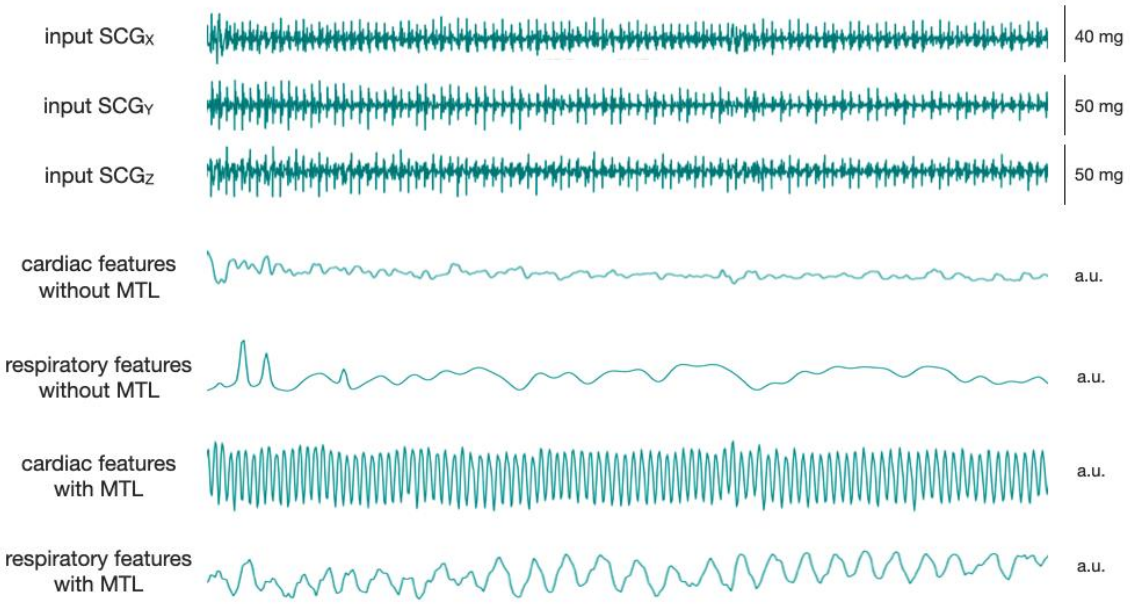


Figure 6.6. Interpretable deep temporal features extracted by $G_{C,SCG}$ and $G_{R,SCG}$. Tri-axial SCG input signals with a duration of 60s were shown on the top for reference. The $t_{C,SCG}$ and $t_{R,SCG}$ features are more interpretable when MTL was applied. The cardiac features extracted by $G_{C,SCG}$ resembled a sinusoidal signal oscillating at the cardiac frequency (101 bpm in this example). The respiratory features extracted by $G_{R,SCG}$ resembled a sinusoidal signal oscillating at the respiratory frequency (30 brpm in this example).

As illustrated in Figure 6.6, $t_{C,SCG}$ and $t_{R,SCG}$ became more interpretable after applying MTL—the cardiac features resembled the smoothened SCG signals while the respiratory features recovered the SDR signals.

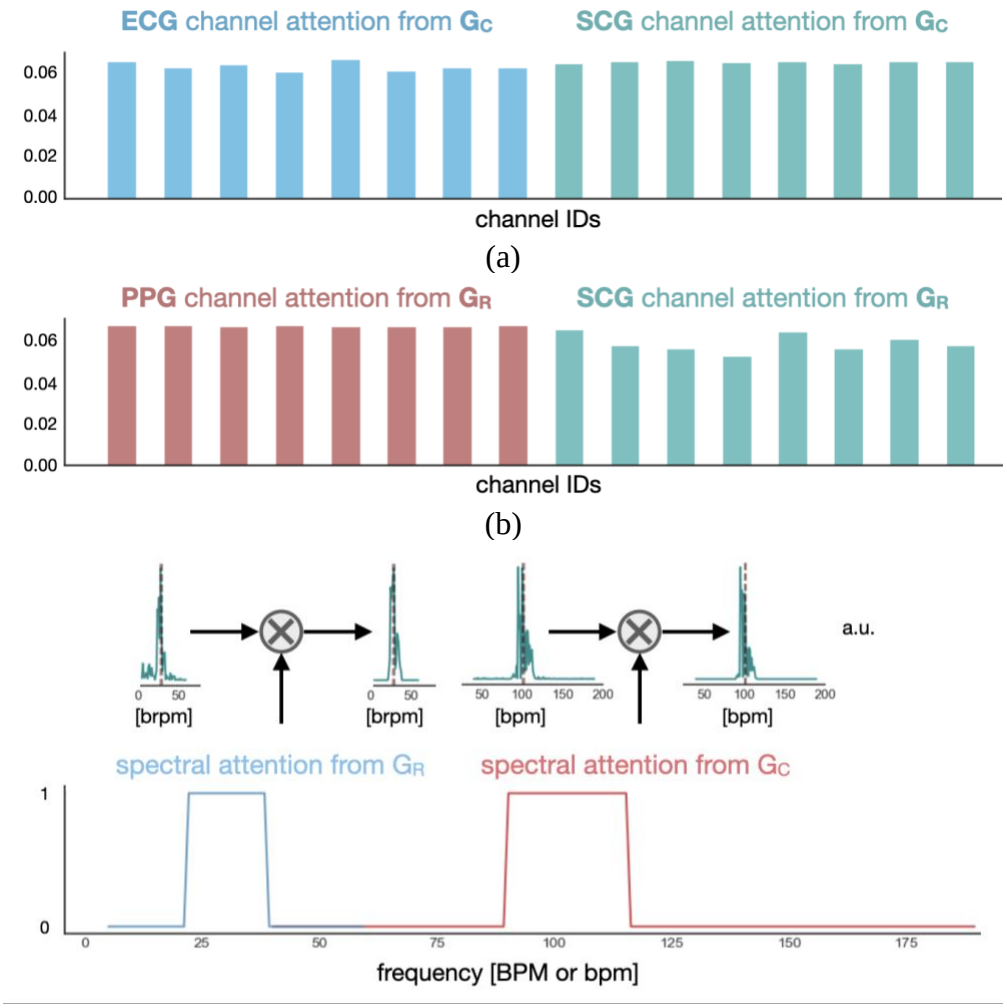


Figure 6.7. Interpretable attentions extracted. (a) Channel attentions of $f_{C,cat}$ derived from ECG and SCG in the HR regression block. (b) Channel attentions of $f_{R,cat}$ derived from PPG and SCG in the RR regression block. (c) Spectral attentions extracted by the RR regression block (shown in blue) and the HR regression block (shown in red). After applying the spectral attentions, out-of-band artifacts were attenuated.

Besides the interpretable deep features extracted for the regression tasks, we showed that both the channel and spectral attentions were also interpretable in Figure 6.7. Again, the channel attention fundamentally represented the quality of the signal of interest. In the example presented, ECG and SCG appeared to have similar cardiac quality, agreeing with the empirical results shown in **CHAPTER 5** while PPG appeared to have stronger respiratory quality than SCG, agreeing with the empirical results shown in **CHAPTER 4**.

The spectral attentions, a_C^f and a_R^f were also in reasonable range. The ground truth HR for this example was around 90 bpm, and a_C^f did pass frequencies from 85-115 bpm. Similarly, the ground truth RR for this example was around 25 brpm, and a_R^f passed frequencies from 23-40 brpm.

6.3.4 Effect of Auxiliary Weights

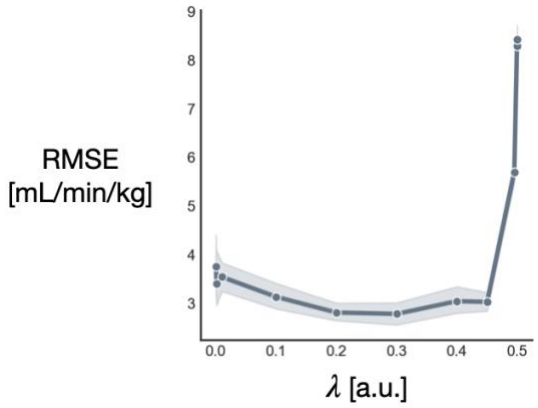


Figure 6.8. Comparison of different λ. The optimal auxiliary weight was around 0.3 (0.4 for the main weight)

As show in Figure 6.8, without auxiliary task ($\lambda = 0$), VO_2 estimation performance was worse. This implied that the auxiliary tasks were indeed helpful to the main task. If the λ was close to 0.5 (main task weight is near 0), the estimation performance was even worse, which would suggest negative transfer (NT) [166] has occurred. The optimal λ was 0.3 in this study.

6.3.5 Selection of Auxiliary Tasks

Table 6.2 Comparison of different auxiliary tasks. Multi-task learning is preferable only when both auxiliary tasks (HR and RR estimation) were used.

	STL	HR only	RR only	HR and RR
RMSE	3.75±0.67	3.20±0.14	3.16±0.17	**2.78±0.16**

To further assess if selecting a certain axillary task may induce NT, we evaluated the VO_2 estimation performance in four scenarios: 1) *Single-Task Learning (STL)*, 2) *HR only* (HR estimation as the auxiliary task only), 3) *RR only* (RR estimation as the auxiliary task only), and 4) *HR and RR* (both HR and RR estimation as the auxiliary tasks). In Table 6.2, *HR only* and *RR only* were slightly better than *STL*. Improvements were demonstrated most in *HR and RR*, suggest z_c and HR_{est} may be complementary to z_R and RR_{est}.

6.3.6 Complementariness of Deep Features and Handcrafted Features

Table 6.3 VO_2 estimation error measured in RMSE using different feature choices and different models.

	regressor	
	DNN	**XGBoost**
z_{ant}	5.13±1.96	4.15±0.01
z_{ant} + z_{hc} [30]	3.00±0.08	2.70±0.06
z_{ant} + z_C + z_R + RR_{est} + HR_{est}	2.89±0.17	2.78±0.16
z_{ant} + z_C + z_R + RR_{est} + HR_{est} + z_{hc}	**2.87±0.15**	**2.44±0.05**

In Table 6.3, we compared the performance with prior work that used z_{hc}. Combining only z_{ant} and deep features extracted from DNN-based CardioRespNet lowered the RMSE when compared to prior work (by -3.67%). If the regressor was replaced with XGBoost, RMSE increased instead (by +2.96%). The lowest RMSE was achieved by merging z_{ant}, z_C, z_R, HR_{est}, RR_{est}, and z_{hc} and estimate VO_2 using XGboost (by -12.23% when compared to no z_{hc}). The results seemed suggest the information carried by the handcrafted features and the deep features may be complementary. When compared to the standard features (i.e., z_{ant}) used to estimate VO_2 for stationary patients, the improvements were very obvious (-41.20%), suggesting anthropometric features should not be used alone if cardiopulmonary information is available. Note that we did not show the results when

using just z_{ant} and HR since they have been shown to be suboptimal to z_{ant} + z_{hc} in prior work [80], [35].

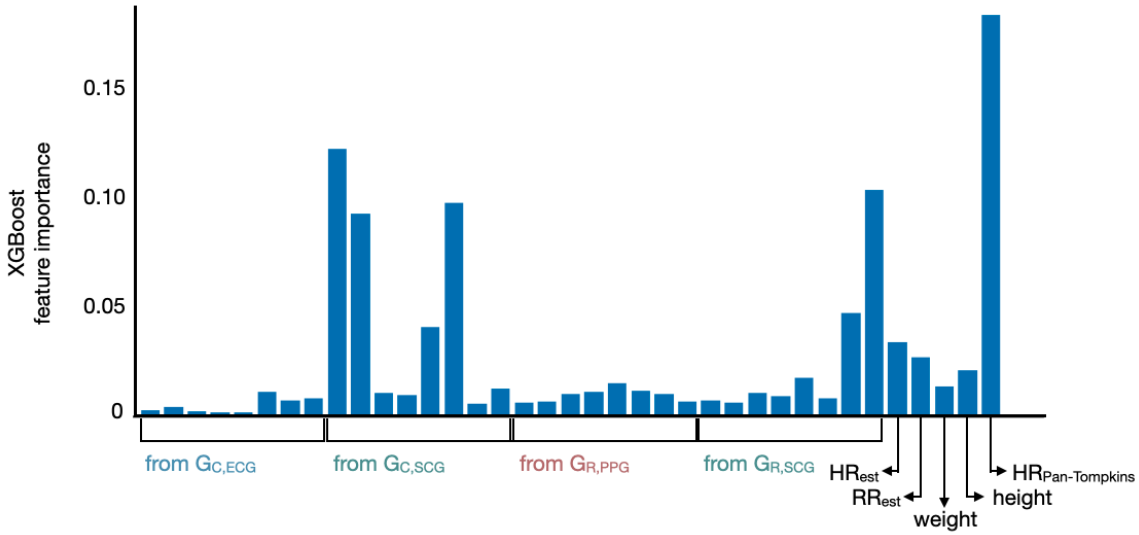


Figure 6.9. XGBoost importance (gain) measured how much accuracy is improved by using a feature. Note that the importance values were normalized so they summed to one.

In Figure 6.9, the XGBoost showed that the most important features include the HR_{est}, RR_{est}, z_{ant}, z_C contributed by x_{SCG}, and z_R contributed by x_{SCG}. On the other hand, z_C contributed by x_{ECG} and z_R contributed by x_{PPG} were not as helpful.

6.4 Discussions

6.4.1 MTL Naturally Decouple Cardiac and Respiratory Information from SCG

The goal of including the auxiliary tasks was to learn representations that can be helpful for the main task [167]. Based on the empirical results, the representations useful for HR and RR were also helpful for VO_2 estimation. Specifically, we believe that $G_{C,SCG}$ may have learned how to smoothen the SCG signals in ways analogous the signal processing procedure for HR estimation designed by Lee *et al.* in [143], which applied

Savitzky-Golay filtering, lowpass filtering, ensemble averaging. Similarly, $G_{R,SCG}$ may have learned how to derive the respiratory signal from SCG in ways analogous to the RR estimation method designed by Chan *et al.* in [18], which detected peaks from smoothened SCG, interpolated them, and finally filtered them to extract SCG-derived respiratory signal. The regularization imposed by the auxiliary tasks enabled the extraction of these features more salient to cardiac and respiratory information.

MTL was particularly suitable for this experimental setup because multiple ground truth labels were collected during data collection. In this work, the ground truth device was the portable gas analyzer COSMED K5 system that recorded VO_2, RR, TV, O_2 fractional concentration in inspired and expired air. In fact, the K5 system directly derived VO_2 using TV, RR, and O_2 fraction inspired and expired air, which explains why learning a good representation of the respiratory features that were related to TV from SCG and RR_{est} from PPG and SCG can be helpful for the main task. On the other hand, the cardiac information was not used by COSMED K5 system to derive VO_2, but Fick's law can still explain their indicative power for VO_2. Although the relationship between CO and VO_2 may not be linear, it was still very well correlated and did not vary too much across subjects, especially below the aerobic threshold [168]. The SV information provided by SCG [16], [169], along with HR_{est} from ECG and SCG, helped provide a surrogate measure of CO. Therefore, it is therefore understandable that SCG cardiac and respiratory features extracted have high feature importance.

Ultimately, maintaining sufficient VO_2 for body's demand would require both the heart and the lung, each provided specialized functions for the transportation of oxygen [5]. Through the both the FT block and the MTL framework, we were able to decouple the cardiac and respiratory information from the SCG measured naturally using the frequency information (i.e., HR and RR). Doing so allowed us to estimate VO_2 more robustly based since we achieved higher accuracy in MTL setting than in STL setting.

6.4.2 Merging Knowledge in Physiology and DSP into CardioRespNet

In this work, for the first time, we merged knowledge in physiology and DSP with a deep learning model and achieved superior performance for VO_2 estimation from non-invasive physiological signals. In the past, DL models designed for such health parameter estimation tasks generally lent architectures that worked in CV, NLP, and ASR because of the plethora of DL approaches in those fields. However, these DL approaches were not specifically designed for such tasks and therefore may not consider alternative, yet reliable assumptions based on the domain knowledge in physiology and DSP. Understandably, as there is not a standardized procedure to transfer such knowledge to a DL model.

By providing the functional meanings of each block of our proposed CardioRespNet, we showed that our model, despite formed using deep convolutional layers, could generate features that met our expectation. Each functional block can be linked to an existing processing procedure in DSP, and the MTL framework allowed us to construct an architecture that respected the relationship between VO_2, HR, and SV, and RR, and TV, which made the design process more logical and debuggable.

6.4.3 Benefits of FT Block

Besides improving model performance, transforming deep temporal features to the spectral domain also has several advantages. First, it simplified the HR and RR estimation task. In temporal domain, estimating the HR and RR may require sophisticated peak detection algorithms that are hard to mirror in a DL model. However, in spectral domain, HR and RR estimation tasks became a simple problem that searched for the dominant frequency. We found that by formulating these tasks as the computation of expected frequency, it is more robust to weight initialization when compared to the argmax function.

After all, if a wrong prediction was made by the argmax function, no gradient was backpropagated and the model did not learn.

Because of the FT block, we can merge DSP methods that were designed for spectral features with CardioRespNet. The channel attention was inspired by the channel fusion used for rPPG signals, which compute the SNR for each channel and a weighted average was used to fuse the multi-channel data. The spectral attention was analogous to a bandpass filter, which first computed the passband and the stopband using the fused deep spectral feature. The fused deep spectral feature was finally filtered using the spectral attention for estimating HR and RR. Since spectral energies within the stopband should not contribute to the estimation of HR and RR, we also believed they provide minimal predictive value of SV and TV. As evident in the results, applying channel and spectral attention also improved VO_2 , HR, and RR estimation performance.

6.4.4 Importance of Deep Features, Estimated HR, and Estimated RR

The spectral features from SCG (i.e., $f_{C,SCG}$ and $f_{R,SCG}$) appeared to be more important than those from ECG and PPG, suggesting the amplitude of the demodulated SCG was more important. Although it has been shown ECG can be used to estimate TV [170], it may be specific to the armband form factor and may not be applicable to the chest-worn form factor. We also noticed that respiratory PPG amplitude can be modulated by TV in a preliminary analysis, but their relationship was generally not subject invariant. This is likely because respiratory PPG can be affected by various factors (e.g., possibly by venous oxygen saturation [171]) even after normalizing using the DC component to account for light strengths, skin color, etc. Finally, to the best of our knowledge, there has been no concrete evidence suggesting SV can be estimate from ECG or chest PPG without calibration. Hence, the feature importance results met our expectation. Finally, we also

demonstrated for the first time, that SCG features representative of respiration were salient for VO_2 estimation.

6.5 Conclusion

This work merged physiology and DSP into an MTL-based DL architecture for extracting deep features that were interpretable and salient for VO_2. The functional meanings of the DL architecture were also provided to explain the logic of the design process and were mirrored to DSP approaches presented in literatures. The channel and spectral attention mechanisms were well suited for deep features transformed to the spectral domain, and the improvements made were not only quantitatively but also qualitatively. Finally, the deep features extracted complemented well to the handcrafted features designed in prior work to estimate VO_2 more accurately. This work will pave the way to further study potential synergistic engineering approaches to combine physiological, DSP, and DL to extract information from physiological signals in a more logical, tractable manner.

From **CHAPTER 2** to **CHAPTER 5**, we have been experimenting with the idea of transforming classical DSP pipelines with ML models slowly and carefully. In **CHAPTER 5**, the DL pipeline was proposed to take filtered signals as the inputs and directly generated the target health parameter as the output. In hindsight, it is hard to tell whether we have progressively developed a DSP pipeline with more DL flavors, or a DL model inspired by DSP modules. This ambiguity may be symbolizing the blurry boundary between DSP and DL, offering an opportunity for the emergence of a feature extraction pipeline that is interpretable, debuggable, yet also capable of learning translational invariant, expressive patterns and being regularized by spare data (i.e., **CHAPTER 5**) and labels (i.e., **CHAPTER 6**). As shown in this chapter, through MTL, we were able to apply

physiological knowledge to design the DSP-inspired DL architecture, one that is aware of the abstract meanings of the features extracted.

CHAPTER 7. CONCLUSION

7.1 Conclusion

The advancements in wearable technologies have enabled the extraction of hidden health information more conveniently and more accurately. These health wearables have already driven peoples' behaviors toward a healthier life, as well as helped discover new knowledge about diseases and our body. The affordable, non-invasive, and physiological sensors used in this work have already realized a health wearable that address the convenience problem well. It is thus our primary goal to further address the accuracy problem in this dissertation. ML fits such problem well owning to its ability to capture the complex relationship between features from physiological signals and the target variables. In literatures, DSP and ML were predominantly stacked sequentially for features extraction and target variable estimation respectively, with physiological knowledge being used during the feature extraction step mostly. Therefore, it was not clear how we could leverage our understanding of the physiological signals, the physiology, and the DSP techniques to benefit ML, beyond simply extracting better features. Through this dissertation, we learned that by designing DL architectures inspired by DSP and physiological knowledge, DL can extract more salient, interpretable features for downstream tasks. Again, the interpretability is loosely defined as the ability to interpret a ML model using engineering concepts. We demonstrated the interpretability of the model through the qualitative depiction of the data that are usually seen in the DSP domain rather than the ML domain. These models are not necessarily interpretable by clinicians' standards.

Again, the main objective of this dissertation was to develop algorithms that can accurately estimate health parameters (i.e., SpO_2, RR, HR, VO_2) from the physiological signals (i.e., ECG, SCG, and PPG). Major breakthroughs have emerged through this effort

because of our approach to marry DSP, physiological knowledge, and ML. The evolution of our approaches from **CHAPTER 2** to **CHAPTER 6** provided examples that showed the possibility to merge seemingly independent engineering approaches and abstract physiological knowledge cohesively. We did not invent new DSP techniques nor novel DL feature extractors that are deeper nor wider. Instead, we proposed using DL in a DSP pipeline and also using DSP and physiological knowledge within DL models to improve feature extraction, demodulation, denoising, multi-model, multi-channel fusion, and regression. This dissertation achieved the following contributions toward developing more accurate methods to estimate health parameters from physiological signals:

1. Adopted classical DSP pipelines and replaced one functional module with data-driven models for the estimation of SpO_2 and RR.
2. Transformed DSP pipelines to achieve feature extraction, fusion, and/or denoising simultaneously using DL models for the estimation of RR and HR.
3. Based on the new DSP-DL hybrid approaches designed in the previous contributions, we designed a DL model that preserved the interpretable, analytical nature of a DSP pipeline and utilized physiological knowledge for the estimation of VO_2.

In the last contribution, DSP modules disguised themselves as a feature projection step, attention mechanisms, and fusion while physiological knowledge disguised itself as regularizers. We were very excited about the research opportunities that these efforts opened and accurate health wearables they may enable. We will discuss some potential research opportunities as future directions in the next section.

7.2 Future Work

First, in this dissertation, the population recruited is generally young and healthy. Although the primary objective of this dissertation is to detect presymptomatic onset of disease infection of seemingly healthy subjects (the diseased population), their data

distributions are not necessarily the same. We leveraged and adapted protocols presented in literatures to induce the physiological changes expected from the diseased population. However, our young and healthy population exhibited physiological changes due to exercise and breath-holds whereas the diseased population may exhibit physiological changes due to more complex interactions of the viruses and the body that are not captured in our datasets. The results presented in this dissertation may serve as strong evidence that the chest-worn wearables could potentially estimate the cardiorespiratory parameters accurately on diseased population, yet further validation of the device accuracy on the diseased population will still be required.

Next, we will consider designing better DL-based feature extractor that considers the inter-channel relationship among the tri-axial SCG signals. Currently, most DL-based feature extractors for 1D physiological signals have relied on temporal convolutional layers for modeling temporal features. The inter-channel relationship of cardiac vibrations that resemble SCG have their own unique characteristics, but no DL methods have been proposed to capture such relationship. By adapting the multi-view convolution presented in [172] for human motion signals, we may be able to develop a DL module to decouple cardiac vibrations from motion artifacts. Motion artifacts presented in SCG might be attenuated through this approach since the inter-channel relationship of SCG could be relatively stable compared that of motion artifacts.

Third, we have only thus far discussed the extraction of amplitude features of SCG in this dissertation, while the cardiac and vascular timing information such as pre-ejection period, left ventricular ejection time, pulse transit time, etc. have shown to provide direct or indirect information of the underlying diseases [33], [46], [57], [153]. The common practices to extract these features were based on DSP techniques, which may be susceptible to overfitting if the DSP pipeline was developed using a small dataset. We can once again leverage the expressive capability of DL to learn features informative of the timing

information. This may be accomplished by extracting the phase information using the FT block (which was used only to extract SCG's amplitude information in **CHAPTER 5** and **CHAPTER 6**). Specifically, timing information could be extracted from the detected "dominant phases" of two signals (e.g., SCG and ECG for pre-ejection period), again, through the expectation operator used in **CHAPTER 5** and **CHAPTER 6**. Moreover, we could also consider replacing the FT block with a wavelet transform block, which may be more useful if both temporal and spectral features are required for estimating the target variable.

Finally, we anticipate the algorithms developed in this dissertation will lead to the discovery of new understandings of physiology and diseases in less controlled, contextualized setting. For example, studying the RR kinetics along with other health parameters during 6MWT for certain disease groups at home has become a more realistic goal owning to the convenient form factor of the chest-worn wearable patch and the accurate RR estimation algorithm presented in **CHAPTER 4**. During 6MWT, are the cardiovascular system or the respiratory system adaptive enough to catch up to the oxygen demand? Is the PRQ within the normal range? If not, does it inform the status of the autonomic nervous system? After 6MWT, how will SpO_2, VO_2, HR, SV, etc. behave? How fast can the body repay "oxygen debt" for patients with heart diseases? Furthermore, if accompanied with other contextual information (e.g., human activity [108], [172], [173], GPS location [174], etc.), can we determine the disease state more precisely? Eventually, we hope to achieve long-term tracking of environmental, activity, and health parameters [81] to study hard yet impactful scientific questions that have been limited by the convenience and the accuracy of the sensing systems. Demonstrating the discoveries of physiological knowledge through this chest-worn wearable patch and its algorithms will signify the revival of the co-evolutionary progression in DSP, physiological knowledge, and DL.

7.3 Final Remarks

During the panel discussion at the Technion-Israel Institute of Technology, Technion Computer Engineering Center (TCE) conference [175], an intriguing question was raised: Is deep learning the final frontier and the end of signal processing? The DSP experts suggested DL was not intelligent as its predictive capability mainly come from the computational and memorization power. We did not know its mechanisms compared to the previous "miracles" such as compressed sensing developed in DSP. For this reason, using the term "deep" to describe such ML model can be misleading. On the other hand, the DL experts reminded everyone that DL approaches are now leading in the field of CV, NPL, and ASR because of their ability to distill information from a large amount of unstructured data. Using DL for such purpose is more reasonable because finding the complex underlying structure of these seemingly unstructured data through domain experts and DSP experts was simply not scalable. I learned and contemplated a lot after hearing this discussion, trying to decide which side is true.

My answer is that they could be both right. The logicality of the DSP pipeline design process was very satisfying to study while at the same time, the expressiveness of DL architectures with diverse utilities also fascinated me. Having been exposed to multiple disciplines ever since I was a biomedical engineering undergraduate student, I learned that the boundaries between disciplines were a lot blurrier than most people think. The rigorous mathematical derivations I did in a Biotransport (for drug deliver) undergraduate course by Dr. Kamei at UCLA has laid a strong foundation to help me grasp the principle of pulse oximetry in **CHAPTER 2** quickly. In a similar vein, having taken both DSP courses and ML courses have made me see very clearly that they do not necessarily conflict with each other or merely stackable. The urge to marry them together was not to reconcile the

conflicts of a controversial topic but rather to use everything to our advantages to address the problems that we believe are impactful. After all, enabling the prevention of diseases, assessment of disease progression, and prediction of the recovery status through wearable technologies have always been our main objective. Only those that are accurate and convenient enough will lead to public adoption and eventually improved clinical outcomes of patients.

REFERENCES

[1] H. Kwon, G. D. Abowd, and T. Plötz, "Complex Deep Neural Networks from Large Scale Virtual IMU Data for Effective Human Activity Recognition Using Wearables," *Sensors*, vol. 21, no. 24, p. 8337, Dec. 2021, doi: 10.3390/s21248337.

[2] H. Kwon, "A Dissertation Presented to The Academic Faculty".

[3] T. PlÖtz, "Applying Machine Learning for Sensor Data Analysis in Interactive Systems: Common Pitfalls of Pragmatic Use and Ways to Avoid Them," *ACM Comput. Surv.*, vol. 54, no. 6, pp. 1–25, Jul. 2021, doi: 10.1145/3459666.

[4] K. Alyafei, R. Ahmed, F. F. Abir, M. E. H. Chowdhury, and K. K. Naji, "A comprehensive review of COVID-19 detection techniques: From laboratory systems to wearable devices," *Computers in Biology and Medicine*, vol. 149, p. 106070, Oct. 2022, doi: 10.1016/j.compbiomed.2022.106070.

[5] A. B. Lumb, "Nunn's applied respiratory physiology." Elsevier/Butterworth Heinemann, Edinburgh] ;, 2005.

[6] G. Quer *et al.*, "Wearable sensor data and self-reported symptoms for COVID-19 detection," *Nature Medicine*, vol. 27, no. 1, pp. 73–77, Jan. 2021, doi: 10.1038/s41591-020-1123-x.

[7] J. R. Halliwill and C. T. Minson, "Effect of hypoxia on arterial baroreflex control of heart rate and muscle sympathetic nerve activity in humans," *Journal of Applied Physiology*, vol. 93, no. 3, pp. 857–864, Sep. 2002, doi: 10.1152/japplphysiol.01103.2001.

[8] J. Whittle *et al.*, "Persistent hypermetabolism and longitudinal energy expenditure in critically ill patients with COVID-19," *Crit Care*, vol. 24, no. 1, p. 581, Dec. 2020, doi: 10.1186/s13054-020-03286-7.

[9] A. Natarajan, H.-W. Su, and C. Heneghan, "Assessment of physiological signs associated with COVID-19 measured using wearable devices," *NPJ Digit Med*, vol. 3, p. 156, Nov. 2020, doi: 10.1038/s41746-020-00363-7.

[10] E. R. Capodilupo and D. J. Miller, "Changes in health promoting behavior during COVID-19 physical distancing: Utilizing Wearable Technology to Examine Trends in Sleep, Activity, and Cardiovascular Indicators of Health," Public and Global Health, preprint, Jun. 2020. doi: 10.1101/2020.06.07.20124685.

[11] A. Lunney, N. R. Cunningham, and M. S. Eastin, "Wearable fitness technology: A structural investigation into acceptance and perceived fitness outcomes," *Computers in Human Behavior*, vol. 65, pp. 114–120, Dec. 2016, doi: 10.1016/j.chb.2016.08.007.

[12] J. Li, Q. Ma, A. HS. Chan, and S. S. Man, “Health monitoring through wearable technologies for older adults: Smart wearables acceptance model,” *Applied Ergonomics*, vol. 75, pp. 162–169, Feb. 2019, doi: 10.1016/j.apergo.2018.10.006.

[13] G. Moody *et al.*, “Clinical Validation of the ECG-Derived Respiration (EDR) Technique,” *Computing in Cardiology*, vol. 13, Jan. 1986.

[14] G. G. Berntson, J. T. Cacioppo, and K. S. Quigley, “Respiratory sinus arrhythmia: autonomic origins, physiological mechanisms, and psychophysiological implications.,” *Psychophysiology*, vol. 30, no. 2, pp. 183–196, Mar. 1993, doi: 10.1111/j.1469-8986.1993.tb01731.x.

[15] B. Semiz *et al.*, “Non-Invasive Wearable Patch Utilizing Seismocardiography for Peri-Operative Use in Surgical Patients,” *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 5, pp. 1572–1582, May 2021, doi: 10.1109/JBHI.2020.3032938.

[16] V. G. Ganti *et al.*, “Wearable Seismocardiography-Based Assessment of Stroke Volume in Congenital Heart Disease,” *J Am Heart Assoc*, vol. 11, no. 18, p. e026067, Sep. 2022, doi: 10.1161/JAHA.122.026067.

[17] K. Pandia, O. T. Inan, G. T. A. Kovacs, and L. Giovangrandi, “Extracting respiratory information from seismocardiogram signals acquired on the chest using a miniature accelerometer,” *Physiol. Meas.*, vol. 33, no. 10, p. 1643, Sep. 2012, doi: 10.1088/0967-3334/33/10/1643.

[18] M. Chan, V. Ganti, and O. Inan, “Respiratory Rate Estimation using U-Net-Based Cascaded Framework from Electrocardiogram and Seismocardiogram Signals,” *IEEE Journal of Biomedical and Health Informatics*, pp. 1–1, 2022, doi: 10.1109/JBHI.2022.3144990.

[19] M. M. Soliman, V. G. Ganti, and O. T. Inan, “Toward Wearable Estimation of Tidal Volume via Electrocardiogram and Seismocardiogram Signals,” *IEEE Sensors J.*, vol. 22, no. 18, pp. 18093–18103, Sep. 2022, doi: 10.1109/JSEN.2022.3196601.

[20] P. H. Charlton *et al.*, “Breathing Rate Estimation From the Electrocardiogram and Photoplethysmogram: A Review,” *IEEE Reviews in Biomedical Engineering*, vol. 11, pp. 2–20, 2018, doi: 10.1109/RBME.2017.2763681.

[21] P. H. King, “Design Of Pulse Oximeters,” *IEEE Engineering in Medicine and Biology Magazine*, vol. 17, pp. 117–117, 1998.

[22] M. Nitzan, I. Faib, and H. Friedman, “Respiration-induced changes in tissue blood volume distal to occluded artery, measured by photoplethysmography,” *J. Biomed. Opt.*, vol. 11, no. 4, p. 040506, 2006, doi: 10.1117/1.2236285.

[23] G. B. Moody, R. G. Mark, A. Zoccola, and S. Mantero, “Derivation of Respiratory Signals from Multi-lead ECGs,” p. 12.

[24] M. Varanini *et al.*, “Adaptive filtering of ECG signal for deriving respiratory activity,” in *[1990] Proceedings Computers in Cardiology*, Sep. 1990, pp. 621–624. doi: 10.1109/CIC.1990.144296.

[25] A. H. Gazi *et al.*, “Robust Estimation of Respiratory Variability Uncovers Correlates of Limbic Brain Activity and Transcutaneous Cervical Vagus Nerve Stimulation in the Context of Traumatic Stress,” *IEEE Transactions on Biomedical Engineering*, pp. 1–1, 2021, doi: 10.1109/TBME.2021.3108135.

[26] C. Rudin, C. Chen, Z. Chen, H. Huang, L. Semenova, and C. Zhong, “Interpretable Machine Learning: Fundamental Principles and 10 Grand Challenges.” arXiv, Jul. 09, 2021. Accessed: Mar. 21, 2023. [Online]. Available: http://arxiv.org/abs/2103.11251

[27] M. A. F. Pimentel, O. C. Redfern, R. Hatch, J. D. Young, L. Tarassenko, and P. J. Watkinson, “Trajectories of vital signs in patients with COVID-19,” *Resuscitation*, vol. 156, pp. 99–106, Nov. 2020, doi: 10.1016/j.resuscitation.2020.09.002.

[28] S. Shah *et al.*, “Novel use of home pulse oximetry monitoring in COVID-19 patients discharged from the emergency department identifies need for hospitalization,” *Acad Emerg Med*, p. 10.1111/acem.14053, Jun. 2020, doi: 10.1111/acem.14053.

[29] S. K. Longmore, G. Y. Lui, G. Naik, P. P. Breen, B. Jalaludin, and G. D. Gargiulo, “A Comparison of Reflective Photoplethysmography for Detection of Heart Rate, Blood Oxygen Saturation, and Respiration Rate at Various Anatomical Locations,” *Sensors*, vol. 19, no. 8, Art. no. 8, Jan. 2019, doi: 10.3390/s19081874.

[30] H. J. Davies, I. Williams, N. S. Peters, and D. P. Mandic, “In-Ear SpO2: A Tool for Wearable, Unobtrusive Monitoring of Core Blood Oxygen Saturation,” *Sensors*, vol. 20, no. 17, Art. no. 17, Jan. 2020, doi: 10.3390/s20174879.

[31] S. Sugino, N. Kanaya, M. Mizuuchi, M. Nakayama, and A. Namiki, “Forehead is as sensitive as finger pulse oximetry during general anesthesia,” *Can J Anesth/J Can Anesth*, vol. 51, no. 5, pp. 432–436, May 2004, doi: 10.1007/BF03018304.

[32] J. Bouten, J. G. Bourgois, and J. Boone, “Hold your breath: peripheral and cerebral oxygenation during dry static apnea,” *Eur J Appl Physiol*, vol. 120, no. 10, pp. 2213–2222, Oct. 2020, doi: 10.1007/s00421-020-04445-y.

[33] O. T. Inan *et al.*, “Novel Wearable Seismocardiography and Machine Learning Algorithms Can Assess Clinical Status of Heart Failure Patients,” *Circ Heart Fail*, vol. 11, no. 1, p. e004313, Jan. 2018, doi: 10.1161/CIRCHEARTFAILURE.117.004313.

[34] V. B. Aydemir *et al.*, “Classification of Decompensated Heart Failure From Clinical and Home Ballistocardiography,” *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 5, pp. 1303–1313, May 2020, doi: 10.1109/TBME.2019.2935619.

[35] M. M. H. Shandhi *et al.*, "Estimation of Instantaneous Oxygen Uptake During Exercise and Daily Activities Using a Wearable Cardio-Electromechanical and Environmental Sensor," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 3, pp. 634–646, Mar. 2021, doi: 10.1109/JBHI.2020.3009903.

[36] C. Schreiner, P. Catherwood, J. Anderson, and J. McLaughlin, "Blood oxygen level measurement with a chest-based Pulse Oximetry prototype system," in *2010 Computing in Cardiology*, Sep. 2010, pp. 537–540.

[37] K. A. *et al.*, "Reflectance Pulse Oximetry for Blood Oxygen Saturation Measurement from Diverse Locations-A Preliminary Analysis," in *2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, Jun. 2018, pp. 1–6. doi: 10.1109/MeMeA.2018.8438781.

[38] M. Kramer, A. Lobbestael, E. Barten, J. Eian, and G. Rausch, "Wearable Pulse Oximetry Measurements on the Torso, Arms, and Legs: A Proof of Concept," *Military Medicine*, vol. 182, no. S1, pp. 92–98, Mar. 2017, doi: 10.7205/MILMED-D-16-00129.

[39] R. Vetter *et al.*, "Frequency Domain SpO2 Estimation Based on Multichannel Photoplethysmographic Measurements at the Sternum," in *World Congress on Medical Physics and Biomedical Engineering, September 7 - 12, 2009, Munich, Germany*, in IFMBE Proceedings, vol. 25/4. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 326–329. doi: 10.1007/978-3-642-03882-2_86.

[40] E. Näslund, L.-G. Lindberg, I. Lund, L. Näslund-Koch, A. Larsson, and R. Frithiof, "Measuring arterial oxygen saturation from an intraosseous photoplethysmographic signal derived from the sternum," *J Clin Monit Comput*, vol. 34, no. 1, pp. 55–62, Feb. 2020, doi: 10.1007/s10877-019-00289-w.

[41] "Pulse Oximeters - Premarket Notification Submissions [510(k)s] Guidance for Industry and Food and Drug Administration Staff," Mar. 2013. http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM081352

[42] M. Arnold, "The surgical anatomy of sternal blood supply," *The Journal of Thoracic and Cardiovascular Surgery*, vol. 64, no. 4, pp. 596–610, Oct. 1972, doi: 10.1016/S0022-5223(19)39718-1.

[43] J. Andersson and E. Schagatay, "Effects of lung volume and involuntary breathing movements on the human diving response," *European Journal of Applied Physiology*, vol. 77, no. 1–2, pp. 19–24, Dec. 1997, doi: 10.1007/s004210050294.

[44] S. Vijayarangan, P. Suresh, P. SP, J. Joseph, and M. Sivaprakasam, "Robust Modelling of Reflectance Pulse Oximetry for SpO2 Estimation," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, Jul. 2020, pp. 374–377. doi: 10.1109/EMBC44109.2020.9176410.

[45] G. Priem, C. Martinez, Q. Bodinier, and G. Carrault, “Clinical grade SpO2 prediction through semi-supervised learning,” in *2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE)*, Oct. 2020, pp. 914–921. doi: 10.1109/BIBE50027.2020.00155.

[46] V. Ganti *et al.*, “Enabling Wearable Pulse Transit Time-Based Blood Pressure Estimation for Medically Underserved Areas and Health Equity: Comprehensive Evaluation Study,” *JMIR mHealth and uHealth*, vol. 9, no. 8, p. e27466, Aug. 2021, doi: 10.2196/27466.

[47] L. M. Reynolds, S. C. Nicolson, J. M. Steven, A. Escobar, M. E. McGonigle, and D. R. Jobes, “Influence of Sensor Site Location on Pulse Oximetry Kinetics in Children:,” *Anesthesia & Analgesia*, vol. 76, no. 4, p. 751???754, Apr. 1993, doi: 10.1213/00000539-199304000-00011.

[48] A. Reisner, P. A. Shaltis, D. McCombie, H. H. Asada, D. S. Warner, and M. A. Warner, “Utility of the Photoplethysmogram in Circulatory Monitoring,” *Anesthesiology*, vol. 108, no. 5, pp. 950–958, May 2008, doi: 10.1097/ALN.0b013e31816c89e1.

[49] J. Pan and W. J. Tompkins, “A Real-Time QRS Detection Algorithm,” *IEEE Transactions on Biomedical Engineering*, vol. BME-32, no. 3, pp. 230–236, Mar. 1985, doi: 10.1109/TBME.1985.325532.

[50] A. P. S. Mohamed Tanveejul *et al.*, “A Study on the Subject and Location Specificity in Reflectance based SpO2 Estimation using R-value based Calibration Curve,” in *2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, Jun. 2020, pp. 1–6. doi: 10.1109/MeMeA49120.2020.9137227.

[51] C. Leys, C. Ley, O. Klein, P. Bernard, and L. Licata, “Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median,” *Journal of Experimental Social Psychology*, vol. 49, no. 4, pp. 764–766, Jul. 2013, doi: 10.1016/j.jesp.2013.03.013.

[52] J. P. A. Andersson, M. H. Linér, E. Rünow, and E. K. A. Schagatay, “Diving response and arterial oxygen saturation during apnea and exercise in breath-hold divers,” *Journal of Applied Physiology*, vol. 93, no. 3, pp. 882–886, Sep. 2002, doi: 10.1152/japplphysiol.00863.2001.

[53] J. Andersson and E. Schagatay, “Arterial oxygen desaturation during apnea in humans.,” *Undersea Hyperb Med*, vol. 25, no. 1, pp. 21–25, Spring 1998.

[54] I. B. Stewart, A. C. Bulmer, J. E. Sharman, and L. Ridgway, “Arterial Oxygen Desaturation Kinetics during Apnea,” *Medicine & Science in Sports & Exercise*, vol. 37, no. 11, pp. 1871–1876, Nov. 2005, doi: 10.1249/01.mss.0000176305.51360.7e.

[55] K. Matsumura, S. Toda, and Y. Kato, “RGB and Near-Infrared Light Reflectance/Transmittance Photoplethysmography for Measuring Heart Rate During

Motion," *IEEE Access*, vol. 8, pp. 80233–80242, 2020, doi: 10.1109/ACCESS.2020.2990438.

[56] V. Krasteva and I. Jekova, "QRS Template Matching for Recognition of Ventricular Ectopic Beats," *Ann Biomed Eng*, vol. 35, no. 12, pp. 2065–2076, Nov. 2007, doi: 10.1007/s10439-007-9368-9.

[57] V. G. Ganti, A. Carek, B. N. Nevius, J. Heller, M. Etemadi, and O. Inan, "Wearable Cuff-less Blood Pressure Estimation at Home via Pulse Transit Time," *IEEE Journal of Biomedical and Health Informatics*, pp. 1–1, 2020, doi: 10.1109/JBHI.2020.3021532.

[58] T. B. Fitzpatrick, "The Validity and Practicality of Sun-Reactive Skin Types I Through VI," *Archives of Dermatology*, vol. 124, no. 6, pp. 869–871, Jun. 1988, doi: 10.1001/archderm.1988.01670060015008.

[59] J. R. Feiner, J. W. Severinghaus, and P. E. Bickler, "Dark Skin Decreases the Accuracy of Pulse Oximeters at Low Oxygen Saturation: The Effects of Oximeter Probe Type and Gender," *Anesthesia & Analgesia*, vol. 105, no. 6, p. S18, Dec. 2007, doi: 10.1213/01.ane.0000285988.35174.d9.

[60] P. E. Bickler, "Effects of Skin Pigmentation on Pulse Oximeter Accuracy at Low Saturation," vol. 102, no. 4, p. 5, 2005.

[61] M. J. Parkes, "Breath-holding and its breakpoint," *Experimental Physiology*, vol. 91, no. 1, pp. 1–15, 2006, doi: 10.1113/expphysiol.2005.031625.

[62] K. L. Glass, T. A. Dillard, Y. Y. Phillips, K. G. Torrington, and J. C. Thompson, "Pulse Oximetry Correction for Smoking Exposure," *Military Medicine*, vol. 161, no. 5, pp. 273–276, May 1996, doi: 10.1093/milmed/161.5.273.

[63] E. D. Chan, M. M. Chan, and M. M. Chan, "Pulse oximetry: Understanding its basic principles facilitates appreciation of its limitations," *Respiratory Medicine*, vol. 107, no. 6, pp. 789–799, Jun. 2013, doi: 10.1016/j.rmed.2013.02.004.

[64] G. Quer *et al.*, "Wearable sensor data and self-reported symptoms for COVID-19 detection," *Nature Medicine*, Oct. 2020, doi: 10.1038/s41591-020-1123-x.

[65] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Advancements in Noncontact, Multiparameter Physiological Measurements Using a Webcam," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 1, pp. 7–11, Jan. 2011, doi: 10.1109/TBME.2010.2086456.

[66] C. Orphanidou, S. Fleming, S. A. Shah, and L. Tarassenko, "Data fusion for estimating respiratory rate from a single-lead ECG," *Biomedical Signal Processing and Control*, vol. 8, no. 1, pp. 98–105, Jan. 2013, doi: 10.1016/j.bspc.2012.06.001.

[67] J. Lázaro *et al.*, "Electrocardiogram Derived Respiratory Rate from QRS Slopes and R-Wave Angle," *Ann Biomed Eng*, vol. 42, no. 10, pp. 2072–2083, Oct. 2014, doi: 10.1007/s10439-014-1073-x.

[68] C. Varon *et al.*, "A Comparative Study of ECG-derived Respiration in Ambulatory Monitoring using the Single-lead ECG," *Scientific Reports*, vol. 10, no. 1, Art. no. 1, Mar. 2020, doi: 10.1038/s41598-020-62624-5.

[69] S. Kontaxis *et al.*, "ECG-Derived Respiratory Rate in Atrial Fibrillation," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 3, pp. 905–914, Mar. 2020, doi: 10.1109/TBME.2019.2923587.

[70] W. Karlen, S. Raman, J. M. Ansermino, and G. A. Dumont, "Multiparameter Respiratory Rate Estimation From the Photoplethysmogram," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 7, pp. 1946–1953, Jul. 2013, doi: 10.1109/TBME.2013.2246160.

[71] M. A. F. Pimentel *et al.*, "Toward a Robust Estimation of Respiratory Rate From Pulse Oximeters," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 8, pp. 1914–1923, Aug. 2017, doi: 10.1109/TBME.2016.2613124.

[72] V. Ravichandran *et al.*, "RespNet: A deep learning model for extraction of respiration from photoplethysmogram," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Jul. 2019, pp. 5556–5559. doi: 10.1109/EMBC.2019.8856301.

[73] T. Choudhary, L. N. Sharma, M. K. Bhuyan, and K. Bora, "Identification of Human Breathing-States Using Cardiac-Vibrational Signal for m-Health Applications," *IEEE Sensors Journal*, vol. 21, no. 3, pp. 3463–3470, Feb. 2021, doi: 10.1109/JSEN.2020.3025384.

[74] N. Alamdari, K. Tavakolian, V. Zakeri, R. Fazel-Rezai, and A. Akhbardeh, "A morphological approach to detect respiratory phases of seismocardiogram," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Aug. 2016, pp. 4272–4275. doi: 10.1109/EMBC.2016.7591671.

[75] 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)*, Osaka: IEEE, Jul. 2013, pp. 4058–4061. doi: 10.1109/EMBC.2013.6610436.

[76] P. H. Charlton, T. Bonnici, L. Tarassenko, D. A. Clifton, R. Beale, and P. J. Watkinson, "An assessment of algorithms to estimate respiratory rate from the electrocardiogram and photoplethysmogram," *Physiol. Meas.*, vol. 37, no. 4, p. 610, Mar. 2016, doi: 10.1088/0967-3334/37/4/610.

[77] C. Orphanidou, “Derivation of respiration rate from ambulatory ECG and PPG using Ensemble Empirical Mode Decomposition: Comparison and fusion,” *Computers in Biology and Medicine*, vol. 81, pp. 45–54, Feb. 2017, doi: 10.1016/j.compbiomed.2016.12.005.

[78] D. A. Birrenkott, M. A. F. Pimentel, P. J. Watkinson, and D. A. Clifton, “A Robust Fusion Model for Estimating Respiratory Rate From Photoplethysmography and Electrocardiography,” *IEEE Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 2033–2041, Sep. 2018, doi: 10.1109/TBME.2017.2778265.

[79] A. Adami, R. Boostani, F. Marzbanrad, and P. H. Charlton, “A New Framework to Estimate Breathing Rate From Electrocardiogram, Photoplethysmogram, and Blood Pressure Signals,” *IEEE Access*, vol. 9, pp. 45832–45844, 2021, doi: 10.1109/ACCESS.2021.3066166.

[80] Md. M. H. Shandhi *et al.*, “Wearable Patch-Based Estimation of Oxygen Uptake and Assessment of Clinical Status during Cardiopulmonary Exercise Testing in Patients With Heart Failure,” *Journal of Cardiac Failure*, vol. 26, no. 11, pp. 948–958, Nov. 2020, doi: 10.1016/j.cardfail.2020.05.014.

[81] M. Etemadi, O. T. Inan, J. A. Heller, S. Hersek, L. Klein, and S. Roy, “A Wearable Patch to Enable Long-Term Monitoring of Environmental, Activity and Hemodynamics Variables,” *IEEE Transactions on Biomedical Circuits and Systems*, vol. 10, no. 2, pp. 280–288, Apr. 2016, doi: 10.1109/TBCAS.2015.2405480.

[82] M. R. Pinsky, “Cardiopulmonary Interactions: Physiologic Basis and Clinical Applications,” *Ann Am Thorac Soc*, vol. 15, no. Suppl 1, pp. S45–S48, Feb. 2018, doi: 10.1513/AnnalsATS.201704-339FR.

[83] C. O’Brien and C. Heneghan, “A comparison of algorithms for estimation of a respiratory signal from the surface electrocardiogram,” *Computers in Biology and Medicine*, vol. 37, no. 3, pp. 305–314, Mar. 2007, doi: 10.1016/j.compbiomed.2006.02.002.

[84] D. A. Tong, K. A. Bartels, and K. S. Honeyager, “Adaptive reduction of motion artifact in the electrocardiogram,” in *Proceedings of the Second Joint 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society] [Engineering in Medicine and Biology*, Oct. 2002, pp. 1403–1404 vol.2. doi: 10.1109/IEMBS.2002.1106451.

[85] M. W. Maciejewski, H. Z. Qui, M. Mobli, and J. C. Hoch, “Nonuniform Sampling and Spectral Aliasing,” *J Magn Reson*, vol. 199, no. 1, pp. 88–93, Jul. 2009, doi: 10.1016/j.jmr.2009.04.006.

[86] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A.

F. Frangi, Eds., in Lecture Notes in Computer Science, vol. 9351. Cham: Springer International Publishing, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.

[87] A. Jansson, E. Humphrey, N. Montecchio, R. Bittner, A. Kumar, and T. Weyde, "Singing voice separation with deep u-net convolutional networks," *18th International Society for Music Information Retrieval Conference*, pp. 23–27, 2017.

[88] I. Perez-Suarez *et al.*, "Accuracy and Precision of the COSMED K5 Portable Analyser," *Front Physiol*, vol. 9, Dec. 2018, doi: 10.3389/fphys.2018.01764.

[89] J. Zia, J. Kimball, S. Hersek, M. M. H. Shandhi, B. Semiz, and O. T. Inan, "A Unified Framework for Quality Indexing and Classification of Seismocardiogram Signals," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 4, pp. 1080–1092, Apr. 2020, doi: 10.1109/JBHI.2019.2931348.

[90] Q. Li and G. D. Clifford, "Dynamic time warping and machine learning for signal quality assessment of pulsatile signals," *Physiol. Meas.*, vol. 33, no. 9, pp. 1491–1501, Sep. 2012, doi: 10.1088/0967-3334/33/9/1491.

[91] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," in *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, in ICML'15. JMLR.org, 2015, pp. 448–456.

[92] M. Yi-de, L. Qing, and Q. Zhi-bai, "Automated image segmentation using improved PCNN model based on cross-entropy," in *Proceedings of 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing, 2004.*, Oct. 2004, pp. 743–746. doi: 10.1109/ISIMP.2004.1434171.

[93] C. H. Sudre, W. Li, T. K. M. Vercauteren, S. Ourselin, and M. J. Cardoso, "Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations," *Deep learning in medical image analysis and multimodal learning for clinical decision support : Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017*, vol. 2017, pp. 240–248, 2017.

[94] O. T. Inan *et al.*, "Ballistocardiography and Seismocardiography: A Review of Recent Advances," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 4, pp. 1414–1427, Jul. 2015, doi: 10.1109/JBHI.2014.2361732.

[95] K. Tavakolian, A. Vaseghi, and B. Kaminska, "Improvement of ballistocardiogram processing by inclusion of respiration information," *Physiol. Meas.*, vol. 29, no. 7, pp. 771–781, Jul. 2008, doi: 10.1088/0967-3334/29/7/006.

[96] H. Nyquist, "Certain Topics in Telegraph Transmission Theory," *Transactions of the American Institute of Electrical Engineers*, vol. 47, no. 2, pp. 617–644, Apr. 1928, doi: 10.1109/T-AIEE.1928.5055024.

[97] C. E. Shannon, “Communication in the Presence of Noise,” *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, Jan. 1949, doi: 10.1109/JRPROC.1949.232969.

[98] K. H. Chon, S. Dash, and K. Ju, “Estimation of Respiratory Rate From Photoplethysmogram Data Using Time–Frequency Spectral Estimation,” *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 8, pp. 2054–2063, Aug. 2009, doi: 10.1109/TBME.2009.2019766.

[99] J. P. Martinez, R. Almeida, S. Olmos, A. P. Rocha, and P. Laguna, “A wavelet-based ECG delineator: evaluation on standard databases,” *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 4, pp. 570–581, Apr. 2004, doi: 10.1109/TBME.2003.821031.

[100] A. M. Yañez *et al.*, “Monitoring Breathing Rate at Home Allows Early Identification of COPD Exacerbations,” *Chest*, vol. 142, no. 6, pp. 1524–1529, Dec. 2012, doi: 10.1378/chest.11-2728.

[101] C. Lakshmipathi, “Respiratory Deterioration and Cardiac Arrest,” *Critical Care Medicine*, vol. 47, no. 1, p. e71, Jan. 2019, doi: 10.1097/CCM.0000000000003433.

[102] E. J. Benjamin *et al.*, “Heart Disease and Stroke Statistics—2018 Update: A Report From the American Heart Association,” *Circulation*, vol. 137, no. 12, Mar. 2018, doi: 10.1161/CIR.0000000000000558.

[103] M. M. Churpek, T. C. Yuen, S. Y. Park, D. O. Meltzer, J. B. Hall, and D. P. Edelson, “Derivation of a cardiac arrest prediction model using ward vital signs*:,” *Critical Care Medicine*, vol. 40, no. 7, pp. 2102–2108, Jul. 2012, doi: 10.1097/CCM.0b013e318250aa5a.

[104] A. Nicolò, C. Massaroni, E. Schena, and M. Sacchetti, “The Importance of Respiratory Rate Monitoring: From Healthcare to Sport and Exercise,” *Sensors*, vol. 20, no. 21, p. 6396, Nov. 2020, doi: 10.3390/s20216396.

[105] M. Chan, V. G. Ganti, J. A. Heller, C. A. Abdallah, M. Etemadi, and O. T. Inan, “Enabling Continuous Wearable Reflectance Pulse Oximetry at the Sternum,” *Biosensors*, vol. 11, no. 12, Art. no. 12, Dec. 2021, doi: 10.3390/bios11120521.

[106] M. Chan *et al.*, “Estimating Heart Rate from Seismocardiogram Signal using a Novel Deep Dominant Frequency Regressor and Domain Adversarial Training,” in *2022 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, Oct. 2022, pp. 158–162. doi: 10.1109/BioCAS54905.2022.9948650.

[107] M. M. Soliman, V. G. Ganti, and O. T. Inan, “Towards Wearable Estimation of Tidal Volume via Electrocardiogram and Seismocardiogram Signals,” *IEEE Sensors Journal*, pp. 1–1, 2022, doi: 10.1109/JSEN.2022.3196601.

[108] S. An *et al.*, “AdaptNet: Human Activity Recognition via Bilateral Domain Adaptation Using Semi-Supervised Deep Translation Networks,” *IEEE Sensors*

Journal, vol. 21, no. 18, pp. 20398–20411, Sep. 2021, doi: 10.1109/JSEN.2021.3095176.

[109] S. An, A. H. Gazi, and O. T. Inan, "DynaLAP: Human Activity Recognition in Fixed Protocols via Semi-Supervised Variational Recurrent Neural Networks with Dynamic Priors," *IEEE Sensors Journal*, pp. 1–1, 2022, doi: 10.1109/JSEN.2022.3194677.

[110] L. M. Nilsson, "Respiration Signals from Photoplethysmography," *Anesthesia & Analgesia*, vol. 117, no. 4, pp. 859–865, Oct. 2013, doi: 10.1213/ANE.0b013e31828098b2.

[111] J. Henricson, J. Glasin, S. Rindebratt, and D. Wilhelms, "Respiratory rate monitoring in healthy volunteers by central photoplethysmography compared to capnography," *Journal of Biophotonics*, vol. 15, no. 4, p. e202100270, 2022, doi: 10.1002/jbio.202100270.

[112] G. de Haan and V. Jeanne, "Robust Pulse Rate From Chrominance-Based rPPG," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 10, pp. 2878–2886, Oct. 2013, doi: 10.1109/TBME.2013.2266196.

[113] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, "Algorithmic Principles of Remote PPG," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 7, pp. 1479–1491, Jul. 2017, doi: 10.1109/TBME.2016.2609282.

[114] P. L. Enright, "The Six-Minute Walk Test," *RESPIRATORY CARE*, vol. 48, no. 8, p. 3, 2003.

[115] K. Tabira, H. Mase, K. Kawahara, and N. Miyamoto, "Validity of chronic obstructive pulmonary disease simulation using straws," *European Respiratory Journal*, vol. 48, no. suppl 60, 2016, doi: 10.1183/13993003.congress-2016.PA2198.

[116] S. Khreis, D. Ge, H. A. Rahman, and G. Carrault, "Breathing Rate Estimation Using Kalman Smoother With Electrocardiogram and Photoplethysmogram," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 3, pp. 893–904, Mar. 2020, doi: 10.1109/TBME.2019.2923448.

[117] H. Jung, J. Kimball, T. Receveur, A. H. Gazi, E. Agdeppa, and O. Inan, "Estimation of Tidal Volume Using Load Cells on a Hospital Bed," *IEEE Journal of Biomedical and Health Informatics*, pp. 1–1, 2022, doi: 10.1109/JBHI.2022.3141209.

[118] J. A. Berkebile, S. Mabrouk, V. Ganti, A. Srivatsa, J. Sanchez-Perez, and O. Inan, "Towards Estimation of Tidal Volume and Respiratory Timings via Wearable-Patch-Based Impedance Pneumography in Ambulatory Settings," *IEEE Transactions on Biomedical Engineering*, pp. 1–1, 2021, doi: 10.1109/TBME.2021.3130540.

[119] W. Wei, K. Vatanparvar, L. Zhu, J. Kuang, and A. Gao, "Remote Photoplethysmography and Heart Rate Estimation by Dynamic Region of Interest

Tracking," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Jul. 2022, pp. 3243–3248. doi: 10.1109/EMBC48229.2022.9871722.

[120] M. Altini and L. E. Dunne, "What's Next For Wearable Sensing?," *IEEE Pervasive Computing*, vol. 20, no. 4, pp. 87–92, Oct. 2021, doi: 10.1109/MPRV.2021.3108377.

[121] A. Sobron, I. Romero, and T. Lopetegi, "Evaluation of methods for estimation of respiratory frequency from the ECG," in *2010 Computing in Cardiology*, Sep. 2010, pp. 513–516.

[122] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-Excitation Networks." arXiv, May 16, 2019. Accessed: Jul. 02, 2022. [Online]. Available: http://arxiv.org/abs/1709.01507

[123] O. Oktay *et al.*, "Attention U-Net: Learning Where to Look for the Pancreas." arXiv, May 20, 2018. Accessed: Jun. 29, 2022. [Online]. Available: http://arxiv.org/abs/1804.03999

[124] M. Rossi *et al.*, "Identification of Characteristic Points in Multivariate Physiological Signals by Sensor Fusion and Multi-Task Deep Networks," *Sensors*, vol. 22, no. 7, Art. no. 7, Jan. 2022, doi: 10.3390/s22072684.

[125] V. Jeyhani, T. Vuorinen, M. Mäntysalo, and A. Vehkaoja, "Comparison of simple algorithms for estimating respiration rate from electrical impedance pneumography signals in wearable devices," *Health Technol.*, vol. 7, no. 1, pp. 21–31, Mar. 2017, doi: 10.1007/s12553-016-0156-0.

[126] C. Qiu, F. Wu, W. Han, and M. R. Yuce, "A Wearable Bioimpedance Chest Patch for Real-Time Ambulatory Respiratory Monitoring," *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 9, pp. 2970–2981, Sep. 2022, doi: 10.1109/TBME.2022.3158544.

[127] N. N. Lepine, T. Tajima, T. Ogasawara, R. Kasahara, and H. Koizumi, "Robust respiration rate estimation using adaptive Kalman filtering with textile ECG sensor and accelerometer," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Aug. 2016, pp. 3797–3800. doi: 10.1109/EMBC.2016.7591555.

[128] D. Liaqat *et al.*, "WearBreathing: Real World Respiratory Rate Monitoring Using Smartwatches," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 3, no. 2, pp. 1–22, Jun. 2019, doi: 10.1145/3328927.

[129] C. Massaroni *et al.*, "Respiratory Monitoring During Physical Activities With a Multi-Sensor Smart Garment and Related Algorithms," *IEEE Sensors Journal*, vol. 20, no. 4, pp. 2173–2180, Feb. 2020, doi: 10.1109/JSEN.2019.2949608.

[130] C. E. Goldfine, F. T. Oshim, S. P. Carreiro, B. P. Chapman, D. Ganesan, and T. Rahman, "Respiratory Rate Monitoring in Clinical Environments with a Contactless Ultra-Wideband Impulse Radar-based Sensor System," *Proc Annu Hawaii Int Conf Syst Sci*, vol. 2020, pp. 3366–3375, 2020.

[131] M. Gwak, K. Vatanparvar, J. Kuang, and A. Gao, "Motion- Based Respiratory Rate Estimation with Motion Artifact Removal Using Video of Face and Upper Body," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Glasgow, Scotland, United Kingdom: IEEE, Jul. 2022, pp. 1961–1967. doi: 10.1109/EMBC48229.2022.9871231.

[132] M.-A. Fiedler, M. Rapczyński, and A. Al-Hamadi, "Fusion-Based Approach for Respiratory Rate Recognition From Facial Video Images," *IEEE Access*, vol. 8, pp. 130036–130047, 2020, doi: 10.1109/ACCESS.2020.3008687.

[133] H. Du, P. J. Newton, Y. Salamonson, V. L. Carrieri-Kohlman, and P. M. Davidson, "A Review of the Six-Minute Walk Test: Its Implication as a Self-Administered Assessment Tool," *European Journal of Cardiovascular Nursing*, vol. 8, no. 1, pp. 2–8, Mar. 2009, doi: 10.1016/j.ejcnurse.2008.07.001.

[134] F. Scholkmann and U. Wolf, "The Pulse-Respiration Quotient: A Powerful but Untapped Parameter for Modern Studies About Human Physiology and Pathophysiology," *Front. Physiol.*, vol. 10, p. 371, Apr. 2019, doi: 10.3389/fphys.2019.00371.

[135] L. Raiano, J. Di Tocco, C. Massaroni, G. Di Pino, E. Schena, and D. Formica, "Respiratory Rate Estimation During Walking/Running Activities Using Principal Components Estimated from Signals Recorded by a Smart Garment Embedding Piezoresistive Sensors," in *2021 IEEE International Workshop on Metrology for Industry 4.0 & IoT (MetroInd4.0&IoT)*, Jun. 2021, pp. 544–549. doi: 10.1109/MetroInd4.0IoT51437.2021.9488448.

[136] M. Chu *et al.*, "Respiration rate and volume measurements using wearable strain sensors," *npj Digital Med*, vol. 2, no. 1, p. 8, Dec. 2019, doi: 10.1038/s41746-019-0083-3.

[137] Y. Ganin *et al.*, "Domain-Adversarial Training of Neural Networks," *arXiv:1505.07818 [cs, stat]*, May 2016, Accessed: Jan. 20, 2020. [Online]. Available: http://arxiv.org/abs/1505.07818

[138] S. Brage *et al.*, "Hierarchy of individual calibration levels for heart rate and accelerometry to measure physical activity," *Journal of Applied Physiology*, vol. 103, no. 2, pp. 682–692, Aug. 2007, doi: 10.1152/japplphysiol.00092.2006.

[139] A. H. Gazi *et al.*, "Digital Cardiovascular Biomarker Responses to Transcutaneous Cervical Vagus Nerve Stimulation: State-Space Modeling, Prediction, and Simulation," *JMIR Mhealth Uhealth*, vol. 8, no. 9, p. e20488, Sep. 2020, doi: 10.2196/20488.

[140] J. L. Mortara, “ECG Acquisition and Signal Processing,” in *Cardiac Safety of Noncardiac Drugs: Practical Guidelines for Clinical Research and Drug Development*, J. Morganroth and I. Gussak, Eds., Totowa, NJ: Humana Press, 2005, pp. 131–145. doi: 10.1007/978-1-59259-884-7_7.

[141] O. T. Inan, D. Park, L. Giovangrandi, and G. T. A. Kovacs, “Noninvasive Measurement of Physiological Signals on a Modified Home Bathroom Scale,” *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 8, pp. 2137–2143, Aug. 2012, doi: 10.1109/TBME.2012.2186809.

[142] H. Jung, J. Kimball, T. Receveur, E. Agdeppa, and O. Inan, “Accurate Ballistocardiogram Based Heart Rate Estimation Using an Array of Load Cells in a Hospital Bed,” *IEEE Journal of Biomedical and Health Informatics*, pp. 1–1, 2021, doi: 10.1109/JBHI.2021.3066885.

[143] H. Lee, H. Lee, and M. Whang, “An Enhanced Method to Estimate Heart Rate from Seismocardiography via Ensemble Averaging of Body Movements at Six Degrees of Freedom,” *Sensors (Basel)*, vol. 18, no. 1, p. 238, Jan. 2018, doi: 10.3390/s18010238.

[144] Y. Zhu, H. Zhang, M. Jayachandran, A. K. Ng, J. Biswas, and Z. Chen, “Ballistocardiography with fiber optic sensor in headrest position: A feasibility study and a new processing algorithm,” in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Jul. 2013, pp. 5203–5206. doi: 10.1109/EMBC.2013.6610721.

[145] H. Lee and M. Whang, “Heart Rate Estimated from Body Movements at Six Degrees of Freedom by Convolutional Neural Networks,” *Sensors (Basel)*, vol. 18, no. 5, p. 1392, May 2018, doi: 10.3390/s18051392.

[146] P. Suresh, N. Narayanan, C. V. Pranav, and V. Vijayaraghavan, “End-to-End Deep Learning for Reliable Cardiac Activity Monitoring using Seismocardiograms,” *arXiv:2010.05662 [cs, eess]*, Oct. 2020, Accessed: Dec. 10, 2020. [Online]. Available: http://arxiv.org/abs/2010.05662

[147] S. M. Pröll, E. Tappeiner, S. Hofbauer, C. Kolbitsch, R. Schubert, and K. D. Fritscher, “Heart rate estimation from ballistocardiographic signals using deep learning,” *Physiol. Meas.*, vol. 42, no. 7, p. 075005, Jul. 2021, doi: 10.1088/1361-6579/ac10aa.

[148] J. Wahlstrom *et al.*, “A Hidden Markov Model for Seismocardiography,” *IEEE Trans. Biomed. Eng.*, vol. 64, no. 10, pp. 2361–2372, Oct. 2017, doi: 10.1109/TBME.2017.2648741.

[149] J. M. Kortelainen and J. Virkkala, “FFT averaging of multichannel BCG signals from bed mattress sensor to improve estimation of heart beat interval,” in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug. 2007, pp. 6685–6688. doi: 10.1109/IEMBS.2007.4353894.

[150] J. Hernandez, Y. Li, J. M. Rehg, and R. W. Picard, "BioGlass: Physiological parameter estimation using a head-mounted wearable device," in *2014 4th International Conference on Wireless Mobile Communication and Healthcare - Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH)*, Nov. 2014, pp. 55–58. doi: 10.1109/MOBIHEALTH.2014.7015908.

[151] C. Szegedy *et al.*, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA: IEEE, Jun. 2015, pp. 1–9. doi: 10.1109/CVPR.2015.7298594.

[152] K.-C. Liu *et al.*, "Domain-Adaptive Fall Detection Using Deep Adversarial Training," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 1243–1251, 2021, doi: 10.1109/TNSRE.2021.3089685.

[153] H. Ashouri and O. T. Inan, "Automatic Detection of Seismocardiogram Sensor Misplacement for Robust Pre-Ejection Period Estimation in Unsupervised Settings," *IEEE Sens J*, vol. 17, no. 12, pp. 3805–3813, Jun. 2017, doi: 10.1109/JSEN.2017.2701349.

[154] X. Chen, J. Cheng, R. Song, Y. Liu, R. Ward, and Z. J. Wang, "Video-Based Heart Rate Measurement: Recent Advances and Future Prospects," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 10, pp. 3600–3615, Oct. 2019, doi: 10.1109/TIM.2018.2879706.

[155] K. Sel, B. Ibrahim, and R. Jafari, "ImpediBands: Body Coupled Bio-Impedance Patches for Physiological Sensing Proof of Concept," *IEEE Trans. Biomed. Circuits Syst.*, vol. 14, no. 4, pp. 757–774, Aug. 2020, doi: 10.1109/TBCAS.2020.2995810.

[156] D. C. Frankenfield, G. Y. Sarson, S. A. Blosser, R. N. Cooney, and J. S. Smith, "Validation of a 5-minute steady state indirect calorimetry protocol for resting energy expenditure in critically ill patients.," *Journal of the American College of Nutrition*, vol. 15, no. 4, pp. 397–402, Aug. 1996, doi: 10.1080/07315724.1996.10718615.

[157] S. Gastinger, A. Sorel, G. Nicolas, A. Gratas-Delamarche, and J. Prioux, "A Comparison Between Ventilation and Heart Rate as Indicator of Oxygen Uptake During Different Intensities of Exercise," *J Sports Sci Med*, vol. 9, no. 1, pp. 110–118, Mar. 2010.

[158] S. Gastinger, H. Sefati, G. Nicolas, A. Sorel, A. Gratas-Delamarche, and J. Prioux, "A new method to estimate energy expenditure from abdominal and rib cage distances," *Eur J Appl Physiol*, vol. 111, no. 11, pp. 2823–2835, Nov. 2011, doi: 10.1007/s00421-011-1900-9.

[159] R. Amelard, E. T. Hedge, and R. L. Hughson, "Temporal convolutional networks predict dynamic oxygen uptake response from wearable sensors across exercise intensities," *npj Digit. Med.*, vol. 4, no. 1, p. 156, Nov. 2021, doi: 10.1038/s41746-021-00531-3.

[160] K. Lu, L. Yang, F. Seoane, F. Abtahi, M. Forsman, and K. Lindecrantz, "Fusion of Heart Rate, Respiration and Motion Measurements from a Wearable Sensor System to Enhance Energy Expenditure Estimation," *Sensors*, vol. 18, no. 9, p. 3092, Sep. 2018, doi: 10.3390/s18093092.

[161] Z. Wang *et al.*, "Enhancing instantaneous oxygen uptake estimation by non-linear model using cardio-pulmonary physiological and motion signals," *Front Physiol*, vol. 13, p. 897412, Aug. 2022, doi: 10.3389/fphys.2022.897412.

[162] J. B. de V. Weir, "New methods for calculating metabolic rate with special reference to protein metabolism," *The Journal of Physiology*, vol. 109, no. 1–2, pp. 1–9, Aug. 1949, doi: 10.1113/jphysiol.1949.sp004363.

[163] S. Gastinger, A. Donnelly, R. Dumond, and J. Prioux, "A Review of the Evidence for the Use of Ventilation as a Surrogate Measure of Energy Expenditure," *JPEN J Parenter Enteral Nutr*, vol. 38, no. 8, pp. 926–938, Nov. 2014, doi: 10.1177/0148607114530432.

[164] J. Bridle, "Training Stochastic Model Recognition Algorithms as Networks can Lead to Maximum Mutual Information Estimation of Parameters," in *Advances in Neural Information Processing Systems*, Morgan-Kaufmann, 1989. Accessed: Apr. 12, 2023. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/1989/hash/0336dcbab05b9d5ad24f4333c7658a0e-Abstract.html

[165] T. Beltrame, R. Amelard, A. Wong, and R. L. Hughson, "Prediction of oxygen uptake dynamics by machine learning analysis of wearable sensors during activities of daily living," *Sci Rep*, vol. 7, no. 1, p. 45738, Apr. 2017, doi: 10.1038/srep45738.

[166] W. Zhang, L. Deng, L. Zhang, and D. Wu, "A Survey on Negative Transfer," *arXiv:2009.00909 [cs, stat]*, Aug. 2021, Accessed: Apr. 26, 2022. [Online]. Available: http://arxiv.org/abs/2009.00909

[167] S. Ruder, "An Overview of Multi-Task Learning in Deep Neural Networks," *arXiv:1706.05098 [cs, stat]*, Jun. 2017, Accessed: Aug. 17, 2021. [Online]. Available: http://arxiv.org/abs/1706.05098

[168] K. C. Beck, L. N. Randolph, K. R. Bailey, C. M. Wood, E. M. Snyder, and B. D. Johnson, "Relationship between cardiac output and oxygen consumption during upright cycle exercise in healthy humans," *Journal of Applied Physiology*, vol. 101, no. 5, pp. 1474–1480, Nov. 2006, doi: 10.1152/japplphysiol.00224.2006.

[169] B. Semiz *et al.*, "Non-Invasive Wearable Patch Utilizing Seismocardiography for Peri-Operative Use in Surgical Patients," *IEEE Journal of Biomedical and Health Informatics*, pp. 1–1, 2020, doi: 10.1109/JBHI.2020.3032938.

[170] J. Lázaro *et al.*, "Electrocardiogram Derived Respiration for Tracking Changes in Tidal Volume from a Wearable Armband*," in *2020 42nd Annual International*

Conference of the IEEE Engineering in Medicine Biology Society (EMBC), Jul. 2020, pp. 596–599. doi: 10.1109/EMBC44109.2020.9175486.

[171] M. Khan *et al.*, "Proof of concept non-invasive estimation of peripheral venous oxygen saturation," *BioMedical Engineering OnLine*, vol. 16, no. 1, p. 60, May 2017, doi: 10.1186/s12938-017-0351-x.

[172] H. Kwon, G. D. Abowd, and T. Plötz, "Complex Deep Neural Networks from Large Scale Virtual IMU Data for Effective Human Activity Recognition Using Wearables," *Sensors*, vol. 21, no. 24, p. 8337, Dec. 2021, doi: 10.3390/s21248337.

[173] H. Kwon *et al.*, "An Explainable Spatial-Temporal Graphical Convolutional Network to Score Freezing of Gait in Parkinsonian Patients," *Sensors*, vol. 23, no. 4, p. 1766, Feb. 2023, doi: 10.3390/s23041766.

[174] M. Altini, P. Casale, J. F. Penders, and O. Amft, "Personalization of Energy Expenditure Estimation in Free Living Using Topic Models," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 5, pp. 1577–1586, Sep. 2015, doi: 10.1109/JBHI.2015.2418256.

[175] *Is Deep Learning the Final Frontier and the End of Signal Processing - Panel Discussion at Technion*, (Jul. 01, 2014). Accessed: Apr. 17, 2023. [Online Video]. Available: https://www.youtube.com/watch?v=LZnAFO5gkOQ