Wearable Health Monitoring: Robust Modeling of Physiology and Behavior from Real-World Sparse-Labeled Sensing

A

Dissertation

Presented to
the faculty of the School of Engineering and Applied Science
University of Virginia

in partial fulfillment
of the requirements for the degree

Doctor of Philosophy

by

Md Ridwanul Alam

December 2020

APPROVAL SHEET

This

Dissertation

is submitted in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

Author: Md Ridwanul Alam

This Dissertation has been read and approved by the examing committee:

Advisor: John Lach

Committee Member: Homa Alemzadeh

Committee Member: John Stankovic

Committee Member: Scott Acton

Committee Member: Laura Barnes

Accepted for the School of Engineering and Applied Science:

Craig H. Benson, School of Engineering and Applied Science

December 2020

Wearable Health Monitoring:
Robust Modeling of Physiology and Behavior from Real-World Sparse-Labeled Sensing

A Dissertation

Presented to

The Faculty of the School of Engineering and Applied Science

University of Virginia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

Md Ridwanul Alam

University of Virginia

Charlottesville, VA

December 2020

Wearable Health Monitoring:
Robust Modeling of Physiology and Behavior from Real-World Sparse-Labeled Sensing

Approved By:

Dr. John Lach (Advisor)
Department of Electrical and Computer Engineering
The George Washington University

Dr. Homa Alemzadeh (Chair)
Department of Electrical and Computer Engineering
University of Virginia

Dr. John Stankovic
Department of Computer Science
University of Virginia

Dr. Scott Acton
Department of Electrical and Computer Engineering
University of Virginia

Dr. Laura Barnes
Department of Engineering Systems and Environment
University of Virginia

Date Approved: September 1, 2020

It is not the knowledge, but the act of learning, that grants the greatest joy.

The never-satisfied man is so strange; if he has completed a structure,

then it is not to dwell in it peacefully, but to begin another.

Carl Friedrich Gauss

To Sophie,

my daughter and my soul,

keep smiling

Acknowledgments

The list of people who must be acknowledged is too long for this section. This interdisciplinary human subject research would not have been possible without the generous supports from National Science Foundation, National Institute of Health, and National Institute of Environmental Health Sciences in the form of grants (IIS-1418622, R01-ES023349, and EEC-1160483), as well as the expertise of my collaborators with engineering and medicine backgrounds from Virginia Tech Carilion Clinic, North Carolina A&T State University, and University of North Carolina - Chapel Hill among others.

My sincere appreciation to my advisor Prof John Lach for securing those funding and hosting those collaborations, and supporting my research over the last five years. Beyond the financial supports and facilities, his mentorship, guidance, and encouragement have not only helped me succeed in my research but also boosted my confidence in pursuing my career goals. I feel fortunate to have him as my mentor, my friend, my guardian, and my idol in navigating the amazing maze of my PhD, especially during the frustrations, struggles, and uncertain times. Besides my advisor, I would like to thank Prof John Stankovic for generously helping me with research guidance and networking. I also appreciate the continued support from the rest of the committee members; Prof Homa Alemzadeh, Prof Scott Acton, and Prof Laura Barnes, along the different steps in this research.

I thank my fellow teammates in the Inertia group at UVA for the wonderful collaborations, the exciting discussions, the sleepless nights, and for all the fun adventures. I learnt a lot from them both while working with them and seeing them accomplish wonders, as well as while mentoring some of them in research and learning about their diverse approach to explore uncharted territories. Many, if not all, successes and failures depicted in this dissertation were achieved or addressed collectively with my teammates. I am grateful to their contributions and feel proud to be a part of this team.

Last but not the least, I would like to thank my family, for standing by me and for my career. I am indebted to my spouse, Noushin, for saving me and the rest of our family from all the mess I make, for being the role model of love and care in our lives, and for continuing to sacrifice her career, her comfort, and her dreams for mine. And to my daughter Sophie, who was born when I started my PhD, and has been sacrificing her childhood since then. She carries the beacon of hope for Noushin and me and has always been the purpose of our pursuit of happiness.

TABLE OF CONTENTS

Abstract

Continuous monitoring of health parameters, especially in real-world out-of-hospital settings, is vital for patients with chronic diseases in preventing acute and hazardous health outcomes. Sensing and machine learning (ML) can facilitate continuous health tracking to reduce such risks. For example, real-time sensing-based prediction of the agitated behavior in dementia patients can prevent harmful escalations, similarly, inference-based monitoring of the respiratory function in asthmatics can prevent asthma attacks.

State-of-the-art sensor-ML researches for health monitoring struggle to transfer in-lab performance to the real-world deployments, as models built with in-lab or limited duration "snapshot" data often fail to generalize temporally and spatially to real-world beyond-training scenarios. But continuous long-term health monitoring data from real-world settings are rare for most health applications and are challenging to acquire due to hurdles in system reliability and data availability, along with user compliance and usability. Challenges in model design with such large data range from inter-disease, inter-patient, and intra-patient variations among the health variables to signal noise and missing data. These challenges are aggravated by the lack of reliable ground truth from real users for such long-term data as well as by the lack of generalizability of the model performance beyond training scenarios.

This dissertation addresses these challenges for wearable sensor-ML systems in the acquisition and utilization of real-world long-term data for human behavior and physiology learning. To achieve robust data collection in real-world settings, a novel wearable-edge platform, named behavioral and environmental sensing and intervention (BESI), is proposed and implemented to facilitate usability, unobtrusiveness, reliability and data availability. Using such real-world data, this work explores learning methods for modeling human health parameters. Toward that goal, novel features are proposed for wearable sensor modalities, namely wrist-worn motion sensors (accelerometer and gyroscope) and chest-worn electrocardiogram (ECG). These features are used in standard instance-based ML methods and sequential models in learning health parameters. To capture the temporal progression of health symptoms, sequential models and temporal ensemble of instance-based models are designed and compared. To overcome the challenges of noisy sparse labels from real-world data, multi-instance learning (MIL) methods are implemented to release the constraint of exact labeling of the whole sequence. Finally, to achieve improved performance and generalizability, this work proposes a novel contextual ensemble learning method, called ConxEns. This method leverages available contextual information in learning 'weak' contextual models, and implement a probabilistic aggregator to infer the health parameters as an ensemble of the inferences of the specialized models. ConxEns is implemented for both classification and regression task and is evaluated for performance improvement, and to demonstrate its potential in generalizing beyond known dataset.

The proposed models and ConxEns pipelines are evaluated in the scope of two major healthcare studies: dementia and asthma care. For the dementia study, the BESI system is evaluated and used to collect patients' wrist motion signals during month-long deployments at their homes. The agitated behavior in dementia patients are modeled as a classification task using the proposed motion features with standard ML models. Multi-instance models MIL-Boost and MI-SVM have demonstrated improved performance compared to single-instance models. Using agitated behavioral symptoms as a context, the proposed ConxEns pipeline is implemented to predict agitation. Result demonstrates robust performance and generalizability. The proposed solutions are applied to another study on asthma care demonstrating the generalizability across health applications. For this study, novel features are proposed for wrist motion and wearable ECG signals and are used to regress the respiratory parameters of study participants using both standard ML models and the proposed ConxEns pipeline. Improved performance and generalization in evaluation demonstrate the robustness of the proposed method for health monitoring. This work leads the way for future research in personalized health modeling toward explainable intervention design.

Chapter 1

Introduction

In the United States, chronic diseases are so prevalent that at least six in ten Americans suffer from one. Chronic conditions, such as cancer, asthma, heart disease, stroke, Alzheimer's disease, diabetes, osteoporosis, or chronic kidney disease, often develop slowly and progress over time with various symptoms and signs. These long-lasting conditions worsen with time and are the leading cause of death and disability among patients, contributing heavily to the nation's $3.5 trillion annual health care costs [1]. For patients suffering from existing chronic diseases, many acute conditions and symptoms pose constant risks of harmful outcomes. Such acute symptoms are often severe and can trigger suddenly for various causes. For example, asthma patients may face an exacerbating attack, osteoporosis patients may suffer from broken bones, heart disease patients are susceptible to heart attacks, and dementia patients demonstrate escalation of agitated symptoms [2]. Multiple physiological and behavioral parameters can be used for assessing and forecasting risks of such acute health crises. Consequently, monitoring such parameters has proven to be vital for patients suffering from chronic health conditions. Such parameters are regularly monitored in clinical settings and are part of early warning systems in hospitals [3]. Out-of-hospital availability of such preventive measures can assist patients and their families at home settings. Along this line, lifestyle behaviors and activities are identified as key factors in preventing and controlling the progression of chronic diseases. For example, tobacco use, poor nutrition, lack of physical activity, and excessive alcohol use are regarded as key lifestyle risks [4]. Monitoring such activity and behavior in day-to-day life can facilitate early intervention to prevent severe outcomes.

Beyond chronic diseases, day-to-day continuous health monitoring is essential for many other health applications. Out-of-hospital health monitoring can prevent hospital readmission for various diseases, prevent pregnancy-related crisis, and track neonatal risk. Similarly, wellness and fitness related applications require continuous monitoring of health-related parameters. Preventing the falling risk of elderly, tracking the eating behavior at home to prevent obesity, monitoring activities of daily living (ADL), and quantifying gait, physical activity, or energy expenditure are some examples of related applications [5-7]. Health applications, such as those mentioned here and more, highlight the need for continuous monitoring of activity, behavior, and physiological parameters in day-to-day life. This dissertation is motivated by such health monitoring applications aiming toward real-world impact in healthcare and wellbeing.

Health monitoring often requires complex instruments or privacy invasive and obtrusive methods, which limits such monitoring to hospitals or care facilities, but not available for daily ubiquitous use [3,5]. Recent advances in sensors and machine learning (ML) are enabling more and more healthcare applications every day. From robotic surgery to automated diagnosis and drug discovery, research in these areas shows the potential for significant impacts in all aspects of healthcare. One such area of active research is wearable sensor-ML systems for out-of-hospital health monitoring.

Advances in wearable sensors and ML methods are unlocking the possibilities of wide applications of such sensor-ML systems. While wearable sensors enable acquisition of various physiological and behavioral signals from on-body locations, they also hold the potential to facilitate unobtrusive and ubiquitous sensing [8-10]. On the other hand, data-driven models and learning approaches can infer various human attributes from those signals and predict many health phenomena. ML models enable capturing even non-linear patterns or high-dimensional clusters from sensor signals in designing complex models of human physiology and behavior. Such ML-powered health monitoring can prevent multitudes of health hazards. These systems can assist patients with chronic diseases by continuous risk tracking, elderly people with remote care toward aging-in-place, help patients after surgery with out-patient monitoring to prevent readmission, and facilitate many similar applications as mentioned before [11-13]. Researches in this area have shown high potential in revolutionizing health monitoring. This dissertation focuses on applications of wearable sensor-ML systems, such as wrist-worn motion and chest-worn ECG sensor-based models of human behavior and physiology, in enabling out-of-hospital, more precisely, residential and free-living health monitoring.

1.1 Challenges in Health Sensing

Researchers are trying out various sensor-ML solutions for various applications. But real-world wide-spread use and acceptance of such systems for health applications still need to go long ways. Among many hurdles, one burning question is the lack of performance of such systems in real-world unknown and unpredictable scenarios. Though such systems perform well in the in-lab controlled settings, these performances often cannot be translated outside those settings [14-16]. Due to this lack of 'realism' for sensor-ML systems, i.e. their ability to operate, as designed and expected, in the real-world unseen, often unknown and unpredictable, scenarios, with real people, in real settings, out-of-hospital health monitoring using wearable sensor-ML systems remains an active area of research.

A major challenge in achieving realism is to make the ML models robust beyond their training data. These models constitute the core of any sensor-ML system, and are trained on input signals or data streams from some specific sources to make decisions about actuation or intervention. While these data-driven models enable sensor-ML systems to learn complex non-linear patterns and relationships, these models often fail to generalize, both temporally and spatially, to unseen data from the real-world. Existing research approaches on sensor-ML systems for health applications often train their models on 'snapshot' data collected from recruited healthy subjects if not on real patients in controlled settings such as labs or clinics. Those models are often validated and tested on data from those same settings. Moreover, the generalization methods and metrics such leave-one-out fails to guarantee performance in real-world unseen scenarios. This challenge emphasizes the need for long-term and continuous data from the real-world uncontrolled settings to use in both training and testing the models and improving generalizability. In this dissertation, I address this model generalization aspect of the challenges toward realism for wearable sensor-ML systems. The associated challenges in this endeavor are nested under two tasks: the acquisition and availability of real-world data and the utilization of such large data.

1.1.1 Acquisition and availability of real-world data

The number of publicly available datasets that contain wearable sensor data are growing. These datasets are acquired for various applications ranging from activity recognition and circadian rhythm learning to gait analysis, elderly fall detection, and energy expenditure quantification. Unfortunately, many of these datasets are limited to application-specific ‘snapshot’ data collected from real patients during any number of sessions often in controlled settings such as clinic visits, in contrast to continuous long-term data from patients’ day-to-day lives. Some other datasets that contain multiple days data either use graduate students or other controlled subjects rather than real patients, or use lab settings such as an IoT instrumented room built for data collection rather than patients’ real residences. Dataset containing real-world long-term data from real patients in their real dwellings for any specific health group is rare if not non-existent.

Multiple challenges associated with long-term real-world wearable sensor data collection systems and processes contribute to this scarcity of related datasets. These challenges range from sustaining the user participation to maintaining the system performance [6,7]. Sustained user participation relies on reduced usability burden. System unobtrusiveness and reduced active interaction help to sustain participation over time. But such properties impose constraints on the wearables power consumption and connectivity modes. Some of the existing researches use wearable devices such as smart watches through companion apps on paired mobile devices such as smart phones, such that watches can continuously stream data via Bluetooth to phones. This approach adds the usability burden of carrying around a paired phone across the residential space along with the compliance requisite of wearing the watch. Such requirement is not always suitable for patient populations of certain age groups or disease severity. Data transmission from wearables to other devices via Wi-Fi consumes higher amount of energy costing the much-needed battery life of the wearable devices in real-world deployments. But, without transmitting data to nearby resourceful devices, real-time processing may cost the wearable device both computational and power resources. Such resources are necessary to meet other usability requirements such as device longevity and ease-of-use.

Moreover, long-term system operation maintenance requires less dependence on human operator involvement and more autonomous solutions, which adds to the power, computing, and connectivity constraints of wearable devices [17-19]. Many existing system monitoring platforms are cloud-based commercial dashboards looking at simple device statistics such as battery life, step counts, etc. Such systems often do not incorporate sensor modality specific error checking or self-sustained recovery mechanisms. Moreover, the metrics used for quantifying such system performance are also less focused on the data availability but more on system operating status. These limitations highlight the need for research in wearable sensing systems platform with a focus on improving data availability.

1.1.2 Model design and evaluation

Learning human health parameters from real-world long-term data faces multitudes of hurdles. Data collected in controlled settings benefit from stationary noise estimation and tangible parameter distributions in modeling application-specific health parameters in a supervised learning approach. On the other hand, long-term data from uncontrolled settings are not only noisy and unbalanced, but also suffer from unreliable ground truth from real users to be used in supervised learning pipelines. The noise in sensor signals originate both from the device hardware and software functionality issues and from the usage characteristics of the patient population. Addressing such noise requires sensor modality-specific algorithm design and implementation. Moreover, curating such data for missing values require context-specific imputation methods, rather than standard single distribution-based methods.

The health application in hand often drives the considerations on model selection and implementation. Some applications require instantaneous health parameter estimation, some highlight the need for capturing temporal patterns of health conditions or behavioral trends, and some require surrogate inference of physiological parameters. Depending on the application, different modeling approaches can be employed. Such specialized models require exhaustive search over the existing methodologies as well as exploring novel methods and modifications to capture application-driven relationships [12]. Standard ML methods rely on hand-crafted features from the sensor signals, where as deep learning approaches train end-to-end models. The challenges of overfitting and lack of generalizability of such methods are greater when the dataset is unbalanced and skewed. As mentioned before, existing approaches such as cross-validation or leave-one-out may not represent the generalizability or the lack of it in trained models depending on the data distribution and curation [16]. While such approaches tend to get very specialized, the challenge that makes it hard is to ensure generalizability of such methods for similar applications.

To learn supervised models from the sensor data collected from real-world settings, active participation from users are required to provide ground truth labels for respective health applications. But users or patients and their caregivers may not always be able to perform this task with accuracy for reasons varying from the disease severity and the burden associated with it to simply lack of skill in technology usage. Consequently, such labels are often temporally imprecise and contain human bias and error in observation. Supervised learning with such imprecise labels may negatively affect the training process, as well as the test outcomes. Learning against such noisy and sparse labels is a major challenge.

Different health applications may have different objectives, some may prefer to achieve higher recall, some want to improve precision, some may prefer to personalize the operating point on the receiver operating curve (ROC). Independent of the objective, improved performance is required for such models to achieve the trust of clinical and medical experts to deploy such systems in patients' residences. Another aspect of useful model design is to emphasize explainability of model outcomes to stakeholders. Such explainability can take various forms, from describing the model prediction mechanism and examining intermediate layer outcomes to explaining the outcomes in certain scenarios especially for erroneous prediction outcomes.

1.2 Related Works

To achieve out-of-hospital health monitoring, various sensor modalities and systems have been explored. For example, residential monitoring systems have evolved over the last two decades with high potential for health applications [6-9]. These systems detect daily activity, significant events like falls, and changes in health status by employing body-worn motion, acoustic, and physiological sensors (heart-rate, skin conductance, etc.), in-home passive infra-red (PIR) sensors, video cameras, pressure pads, and smart assisted living systems with multiple modalities. Such monitoring systems often use single to multiple sensing modalities, and only a handful integrate wearables into in-home sensor systems [8,9].

Some related works focus only on in-situ sensor-based systems. For example, Aware Home [20] was proposed as an intelligent space for inferring ADL of inhabitants using vision and acoustic tracking with other sensing, and functionality was prioritized over privacy issues. House_n [21] installed numerous sensors in every part of a residence while acknowledging requirements of reliability and robustness in deployments for behavior inference. Many other similar facilities have been designed and built to evaluate sensing functionalities and validate their usability for health monitoring applications. But such settings remain controlled and fail to generalize to patients' free-living habitats.

To explore real-world scenarios, some works populate patients' residences with multiple sensing modalities. BehaviorScope used cameras, PIRs, and door sensors to estimate activity patterns by detecting indoor location which may not be reliable in multi-resident and high social density spaces [22]. CareWatch was designed to assist caregivers of people with dementia by tracking their location only at night using a commercial security system and a bed occupancy detector [23]. At-Ease used an application specific system of PIRs and water sensors using a single communication protocol, but emphasized challenges of real deployments, reliability, and appearance [24]. UUTE monitored health status and sleep time of users and presented learned lessons from their trial on required unobtrusiveness and reliability of the system [25]. Privacy and security of invasive modalities such as camera remain a major concern for such systems. These ubiquitous sensing systems have demonstrated better performance in other applications such as household energy consumption or remote home security than health applications. While in-situ sensors are good at capturing context, those modalities often fail to sense or infer physiological and behavioral parameters in an instantaneous fine-grained scale.

Adding the benefit of mobility and unobtrusive physical contact of wearable sensors, few in-situ sensing systems are developed with wearable components. ALARM-NET extensively used PIR sensors in living space and a TinyOS based wearable mote to estimate ADL, which was dependent on back-end processing for real time performance [26]. Tapia et al. proposed a multi-agent architecture for ambient intelligence to incorporate autonomy in telemonitoring, but the reliability was dependent on a centralized agent [27]. Dem@care extensively uses video cameras and instrumentation of household objects with other sensors to track activities of users [28], but has not addressed issues like privacy concerns and obtrusiveness. These systems validate the sensing platforms in controlled settings, but lack the ability to incorporate ML-based real-time health monitoring capabilities.

State-of-the-art technologies for the Internet-of-things (IoT) and smart homes incorporate remote monitoring functionality. MavHome [29] employed data mining, prediction, and multi-agent systems to construct models of inhabitant activities from installed sensors and predict possible crises, with lesser focus on unobtrusiveness and the reliability of the architecture. Lotfi et al. presented a network of commercial in-situ sensors to infer location of users without interfering with regular activities [30]. But their system is restricted to single occupant homes and provides no wearable support. HomeOS [31] also presents a platform for commercial emplaced sensors, but the number of allowed vendors is limited. Viani et al. proposed an architecture which was confined to location inference and assumed placement of wireless access points all over the residence without fault handling considerations [32].

On the other end of the spectrum, various wearable and mobile systems have been proposed for out-of-hospital health monitoring. Elaborative descriptions of early wearable systems are presented in [9]. Early systems like LifeGuard [33] and LiveNet [34] were designed to sense various physiological parameters and to transmit to a central hub for real-time processing. A multi-patient monitoring platform named CodeBlue [35] addressed the need for reliability in realistic usage. AMON [36] emphasized reducing the size of wearables and proposed a wrist-worn device to measure physiological indicators and to transmit data using a GSM cellular link. Often the wearable transmitted data to a nearby receiver station (e.g., FireLine) or received data from emplaced transmitters (e.g., Listense) [37]. A similar approach was adopted by Escort system [38], where optical location signal transmitted by Talking Lights were received by the wearable device and transmitted to a central server for location estimation. Such designs impose restriction on range and availability of services, and continuous transmission poses a challenge on power consumption and battery life. The KITE project [39] introduced user participation in the platform design process to avoid stigmatization and increase user compliance. Rapid development of wearables is motivated by the requirements for advanced sensing as well as transparency and unobtrusiveness.

The last decade has seen the advent of smartphones as a novel platform for various health applications [40]. Mobile technologies can passively collect information as a way to reduce burden and improve care for healthcare consumers. Smartphones facilitate passive sensing with multiple housed sensors and their increasing computational power and pervasiveness. Moreover, behavioral data such as social interaction and social media activity can be measured using smartphones. Such sensing has been used in studies on bipolar disorder, depression, schizophrenia, stress, addiction, medication adherence, and other physical and mental health applications [41-43]. Beyond patient-focused health applications, sleep quality assessment, sleep-wake cycle monitoring, energy expenditure, calorie loss, smoking behavior monitoring, and many similar projects are proposed [44-45]. Such studies require patients to always carry the phone, sometimes even to strap the phone on arms. Privacy and security have been major concerns as such approaches look into people's social media and interaction activities. Lack of physiological sensing capabilities and the unrealistic assumption of a phone being used or carried constantly have remained as major concerns with such platforms.

Recent advances in sensing technologies facilitate many forms of wearable sensors, from sensors that measure motion on the wrist to sensing electrocardiogram on the chest, as wrist bands to skin patches, even as commercial products such as smart watches, shoes, textiles, etc. Among those, smart watches have gained popularity and some level of social acceptance both for fashion trend as well as for fitness and well-being awareness [46-48]. Such wrist-worn devices are relatively comfortable to wear and less burdensome to use compared to other means of wearables. They provide many utilities such as texting, email, fitness tracking, etc. on the wrist to increase user engagement. Recent years, smartwatch and wearable platforms have gained attention and popularity in health monitoring. Such platforms have physical contact with human body and can facilitate physiological sensing along with physical activity and behavioral signals. Wearable sensors can measure pulse rate, electrocardiogram (ECG), photoplethysmogram (PPG), blood oxygen levels, skin conductance, electrodermal activity (EDA), bioimpedance, galvanic skin response (GSR), as well as motion through accelerometers, gyroscopes, and magnetometers. These sensing capabilities have shown promise in various health applications. For example, physical activity monitoring has proven useful for multiple patient populations including arthritis [49,50], lumbar fusion surgery [51], arthroplasty surgery [52], and heart failure patients [53]. The association between physical activity and health issues such as blood glucose, lipid profile, renal function, blood pressure, and weight are also studied [54]. Accelerometer and gyroscope are used in many researches for fall detection in elderly care, as well as monitoring their ADL [55]. Geriatric depression and stress tracking solutions are proposed using motion, EDA, and PPG signals. Similar methods are used for gait analysis, bradykinesia, tremor, and dyskinesia quantification in Parkinson's disease, as well as for patients with essential tremor [56-58]. Beyond fitness and wellbeing [59], researches use wearables for stress monitoring [60], and energy and calorie expenditure tracking [61]. Sleep monitoring and sleep quality assessment have been extensively used for various health applications [62, 63]. Researches have used wearables to collect sleep information, evaluate multi-construct cognitive efficiency, and predict changes to mental acuity [64]. Activity, heart rate, and sleep data can be used to monitor pregnancy health [65,66,67]. In epilepsy patients, motion signals are used to identify tonic–clonic seizures [68]. Wearables are widely explored for monitoring smoking behavior [69], tracking and reducing sedentary behavior, for example for patients with obstructive pulmonary diseases [70]. Studies examine the effect of using health monitoring on kids with acute lymphoblastic leukemia and identify the relationship between steps per day and fatigue [71]. Smart wearables are employed to collect vital signals such as EDA and respiration to estimate physical fatigue perception [72]. Scarce processing resources, limited battery life, sensing accuracy, and connectivity issues pose major challenges for translating such researches to real-world settings [73-75].

In addition to the sensing platforms, advances in ML methods, especially deep learning-based architectures, have enabled complex, even non-linear, health parameter and relationship learning [76]. Most of the above-mentioned applications of wearable sensors use some form of machine learning if not deep learning models. For example, convolutional neural network (CNN), recurrent neural network (RNN), long short-term memory (LSTM), auto-encoders, and attention mechanisms in deep networks are used to model ADL from public datasets [77,78]. Such applications often utilize defined patterns for the motion events to model the activity of interest. RNN models are proposed to model fall detection as a sequential activity [79]. LSTMs with CNN have shown promise in stereotypical motor movement inference in autism [80]. Variants of LSTM are applied and evaluated in sleep monitoring and assessment applications [81]. Customized CNN architectures tend to capture important patterns from PPG signals to detect atrial fibrillation [82]. CNN and RNN are explored to model fall detection in elderly population, as well as dementia patients wandering risk monitoring [83]. Among other applications, autoencoders, deep belief networks, and deep Boltzmann machines are used for emotion recognition, epileptic seizure detection, heart rate monitoring, and blood pressure tracking [76, 84]. Variants of deep neural networks have enabled the modeling of various physiological and behavioral patterns. These models, trained and tested on ‘snapshot’ datasets, sometimes perform notably better than traditional instance-based learning methods in controlled datasets. But such performances do not guarantee robust performance in real-world day-to-day settings. Moreover, the lack of explainability is another major drawback for such methods, especially in the context of health applications.

Many of the above depicted systems and methods are application-specific and evaluated on offline datasets acquired in controlled settings or on timeframes of real-world scenarios. Unfortunately, such approaches do not generalize well to times and spaces that are not captured by the training data [12,15]. Baig et al. surveyed smart components of available remote, mobile, and wearable monitoring systems and emphasized the need for reliability, unobtrusiveness, and privacy in such systems [74]. These challenges of developing general purpose systems are also highlighted by Alemdar and Ersoy [6]. These works suggest the need for robust, privacy-preserving, error resilient, and unobtrusive sensor-ML platforms for out-of-hospital health monitoring applications.

1.3 Research Scope and Objectives

With the motivation toward real-world out-of-hospital health monitoring, in this research, I address the challenges for wearable sensor-ML systems in the acquisition and utilization of real-world long-term data for human behavior and physiology modeling. The scope of this research is kept limited to wearable sensors in real-world residential and free-living settings; wrist-worn accelerometer and gyroscope and chest-worn ambulatory ECG are used in this work. Using these modalities as the wearable component, I propose a novel system architecture and demonstrate its ability to overcome the challenges in real-world month-long sensor data acquisition from real patients in their residences. To address the challenges of model design and evaluation with real-world long-term sensor data, this work focuses on two major health application: dementia care and asthma care. Using these application contexts, I propose novel methodologies to discover biomarkers and to model application-specific physiological and behavioral health parameters.

In addressing these broad challenges of real-world long-term data acquisition and utilization, within the scope of the sensing modalities and the specific health monitoring applications, the objective of this dissertation is to answer three research questions (RQ):

RQ-1: *What causes data loss during long-term real-world deployments of wearable sensors in residential settings and how to reduce such loss under usability and resource constraints?*

RQ-2: *How to learn from long-term sensor data with noisy and sparse labels in building application-driven models of human behavior and physiology?*

RQ-3: *What methods can improve the performance and the generalizability of learnt health parameter models with supplemental data labels?*

To answer the first question, I explore wearable sensing challenges in real patients' residences as part of the dementia care study. To overcome the challenges of usability and system reliability, I propose a novel wearable-edge interconnected ML platform. This platform, named behavioral and environmental sensing and intervention (BESI), leverages in-home network of edge devices to alleviate the computing and connectivity constraints from the wearable devices, as well as to facilitate hierarchical autonomous monitoring of system performance over time [85,86]. This architecture reduces the usability burden of the wearables by alleviating constraints to carry a companion phone, while enabling both temporal and spatial continuity in the home setting for months. Moreover, it is equipped with hierarchical watchdogs, heartbeat messaging, and autonomous fault tolerant and recovery methods to enable sustained long-term performance. The BESI system has been deployed in multiple dementia patient residences spanning a total duration over 20 months, and has acquired long-term continuous wrist-motion data of about 100 GB, thus demonstrating robust data acquisition in real-world settings.

The second and third research questions are dependent on respective health monitoring applications and are answered for the dementia and asthma care applications independently with a goal to explore the value of long-term continuous wearable sensor data in modeling human behavior and physiology. For the dementia care study, I model and monitor the physical agitation of dementia patients from their wrist motion data addressing real-world label-noise and missing label issues associated with labels acquired from real users. Beyond exploring learning methods to overcome sparse label issues, I also explore various aspects of application-specific model design and address related challenges. For example, to model the progression of the agitated behavior in dementia patients, I evaluate sequential modeling approaches. And, for the asthma study, the objective is to infer the respiratory parameters from participants motion and wearable ECG signals. In this application, the instantaneous physiological relationship between respiratory parameters and motion and ECG signals is more emphasized than temporal correlation.

In answering the third research question, I propose a novel contextual ensemble pipeline to learn behavioral and physiological parameters from real-world wearable sensor signals with supplementary information beyond exact classification or regression labels. This method has demonstrated performance improvement for both agitation prediction and respiratory parameter inference, highlighting generalizability. Moreover, the hierarchical organization of the pipeline is used to explain model outcome from health perspectives. This focused work captures the trend of many health monitoring applications and can be easily adapted by many.

In search of the answers to the above research questions, I establish with evidence my thesis:

A holistic approach from fault tolerance in sensor systems to

contextual association of data-driven models improves

the robustness of sensing-driven models of

human behavior and physiology in real-world settings.

The thesis in this research is established by pursuing two goals: long-term data acquisition from the real-world and using that data for achieving improved performance, explainability, and generalizability of real-application focused ML models against sparse noisy labels. Toward those goals, I built the BESI system with robustness and reliability features, evaluated and deployed in real dementia patient residences for years, and successfully acquired long-term continuous wearable sensor data. In utilizing those data, I designed and evaluated supervised learning methods for predicting physical agitation of dementia patients from their wrist motion signals and for inferring breathing rate and minute ventilation from wrist motion and wearable ECG signals. The performances of model-based predictions are improved by addressing the imprecision inherent in ground truths or labels provided by real users. Moreover, I propose a novel contextual ensemble method to utilize secondary information associated with the primary data labels to improve predictive performance as well as demonstrating generalizability.

The contributions of this dissertation can be summarized as:

A1. A novel and robust wearable-edge ML platform for long-term behavioral and physiological sensing and inference toward real-world out-of-hospital health monitoring applications,

A2. A baseline ML model for predicting physical agitation in dementia patients from their wrist motion signals with biomarker discovery and supervised learning,

A3. A baseline inference methodology for respiratory parameter estimation from wrist motion and wearable ECG signals across various physical activity context,

A4. A novel supervised learning methodology to learn from long-term real-world sensor data with scarce and imprecise, but multidimensional, labels.

In this dissertation, I present the conducted research on the system design and the ML methodology along with the respective implementations, evaluations, and limitations. In Chapter 2, the healthcare applications on dementia and asthma care are presented in details along with the description of the data collected from these studies. The BESI system is described with its design considerations and robustness features in Chapter 3, and its performance in acquiring long-term continuous data from the real-world. Chapter 4 introduces the sensor biomarker discovery pipeline for both the applications. The agitated behavior and the respiratory physiological parameters learning pipelines are presented in Chapter 5. Chapter 6 presents the results of these methodologies and discusses the performance, explainability, and generalizability. I conclude this dissertation in Chapter 7 with discussions on the achievements, the limitations, and the potentials of this research in advancing health monitoring-based real-time care delivery.

Chapter 2

Studies on Health Monitoring

Two real-world health monitoring applications, one on dementia care, the other on asthma care, directly motivate and contribute as the sources of data and the testbeds for the performance evaluation of this research. In this chapter, I introduce these projects along with their respective goals and challenges. Moreover, the data originated from these projects are also described along with the sensors and media used to acquire those.

2.1 Monitoring Agitated Behavior in Dementia

Dementia is a neurodegenerative chronic disease, about 5.6 million people in US are struggling with Alzheimer's dementia. Dementia patients suffer from memory loss, cognitive, visual, and vocal impairments. These symptoms progressively worsen over time and patients often express agitated behaviors. 90% of dementia patients demonstrate some form of agitation. Agitation in dementia is described as a set of behaviors often repetitious, socially inappropriate, and aggressive in nature. According to the Cohen-Mansfield Agitation Inventory, these behavioral symptoms include wandering, restlessness, feeling lost, inappropriate dressing, throwing up food, kicking, biting, breaking things, lack of attention, repetition, mood swings, imitating, whining, and cursing (Fig 2.1). These symptoms are categorized as physically non-aggressive, physically aggressive, verbally non-aggressive, and verbally aggressive behaviors [87-89]. Such symptoms are harmful both physically and mentally for both the patients and their family caregivers. Hence, caring for a person with dementia (PWD) can be physically and emotionally taxing for caregivers.

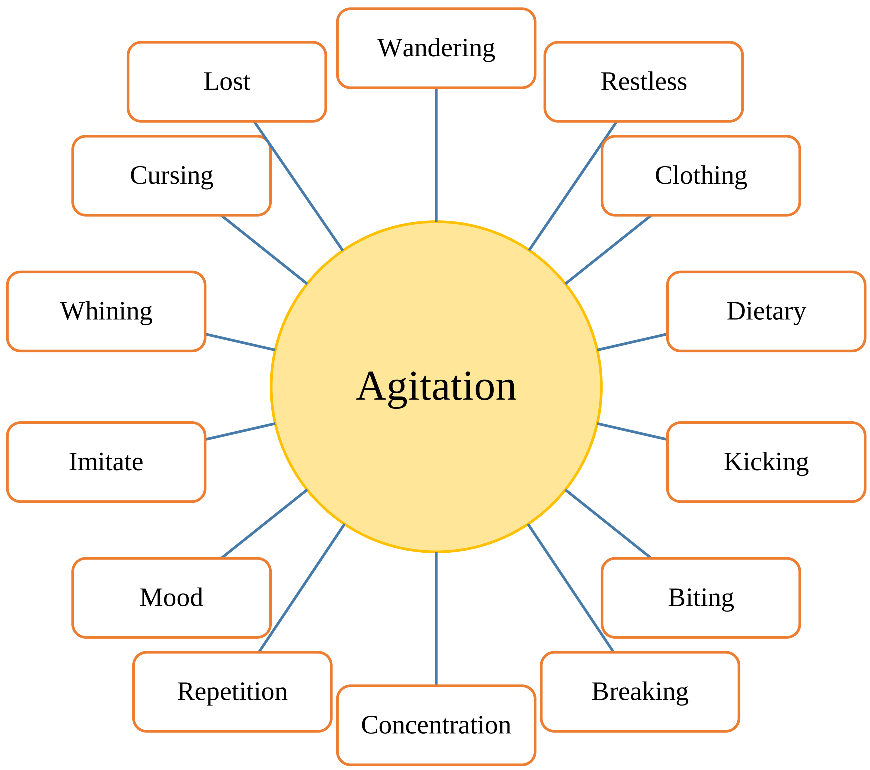


Figure 2.1: Symptoms of agitated behavior in dementia patients according to CMAI.

The dementia care study, named behavioral and environmental sensing and intervention (BESI) for dementia caregiver empowerment, aims to detect and predict agitation in an early stage. Such early detection can enable early notification to the caregiver with personalized interventions, and thus, prevent escalation of agitated behavior. BESI is a three-phase, five-year long study funded by the NSF under the Smart & Connected Health program. The overall goal of the study is to empower caregivers, reduce caregiving burden, and increase self-efficacy by providing just-in-time interventions. From an engineering research perspective, the objective is to continuously monitor and generate model-based predictions on the agitated behavior of dementia patients at their residences (Fig 2.2). Continuous monitoring of such behavior at home not only can enable preparedness and self-efficacy in caregiving but also facilitate longitudinal study of disease progression and intervention design.

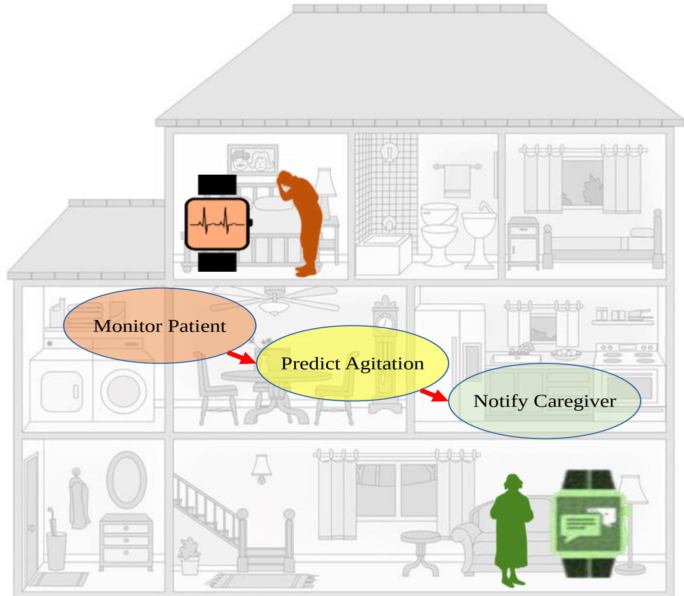


Figure 2.2: The BESI study aims to facilitate just-in-time intervention through continuous behavior monitoring

Existing approaches are often reactive and are administered too late in an escalation to be routinely effective. Standard clinical tools such as the clinical dementia rating (CDR), the Cohen-Mansfield agitation inventory (CMAI), and the modified mini-mental state (3MS) examination are often used for quantitative assessment of the behavioral and psychological symptoms of dementia (BPSD) including agitation [89]. These tools attempt to evaluate the patients' cognitive and functional performance as well as aggressive and non-aggressive behaviors, but are prone to subjective bias and recollection errors from caregivers or clinicians during clinic visits or interview sessions.

For proactive interventions, researchers recently emphasize continuous behavior tracking approaches based on sensing technologies. State-of-the-art sensing-based dementia care mostly focus on monitoring daily activity, detection of early dementia, and quantitative assessment of therapeutic intervention for cognitive care [90-94]. For example, subjective clinical scores are correlated with daily activity levels from wearable sensor data to supplement dementia diagnosis. Similarly, wearable sensors have been used to quantify the treatment outcomes for dementia patients with agitation symptoms, and to evaluate the prospect of early diagnosis through detecting changes in circadian behavior. However, continuous and real-time behavior sensing and prediction is still an active area of research.

In BESI, I use an accelerometer sensor housed in a smartwatch to sense the wrist motion of the dementia patients [95-100]. My research objective is to continuously monitor and to learn and predict the physical agitation behavior of patients. The study population is comprised of 12 patient-caregiver dyads. The patient population consists of 7 women and 5 men, who have been diagnosed with mild to severe dementia. The subjects are recruited often through dementia support groups and community advocacy groups. The demographic of this population is presented in Table 2.1. The patient-caregiver dyads provide consents to participate in the study following a protocol approved by the IRB at Virginia Tech Carilion School of Medicine. Following the protocol, data are collected from the patients at their residences for a continuous duration of 30 to 60 days.

The inertial accelerometer sensor on the Pebble smartwatch, which was worn by the dementia patients during the month-long deployments at their residences, was sampled at 50 Hz to continuously collect the 3-axes motion signal. A custom app named Pixie was designed to sample the accelerometer and to send that data over the sensor network. This accelerometer was the sole modality used in this study to continuously monitor the patients' behavior.

The caregivers in this study also wore a Pebble smartwatch, which was programmed to host an event-marking app named Memento. This app enabled the caregivers to timely mark sparse and unpredictable behavioral episodes like agitation. The timestamp of the button-press is stored as an annotation for an agitation episode on-board, as well as transmitted to a tablet. The tablet hosts an Android app, which the caregivers were requested to use for providing temporal, spatial, and characteristic observations or symptoms (e.g. pacing, hitting, repetition, shouting, etc.) about the agitation episodes, as well as the caregivers perceived severity level of the episode. The timestamps and these supplementary

TABLE 2.1 DEMOGRAPHIC COMPOSITION OF THE BESI STUDY

Sex	Age	CDR (range 0:3)	3MS (range 0:100)	CMAI-C (range 0:203)
F (7)	80.6±4.3	1.55±0.4	59±11	64±11
M (5)	78.7±6.8	1.4±0.27	52±18	60±14

information about the agitation episodes are the ground truth labels and associated context for learning the agitated behavior from the accelerometer signal in a supervised manner.

Over the course of about 20 months, the caregivers marked 571 agitation episodes for those twelve patients. Since an agitation episode lasts about 20 minutes on average, there were 11,420 minutes of data for agitation episodes, which represents 34,260,000 x 3 samples of sensor signals. Though this number seems very large, considering that the raw sensor signal samples do not carry much information independently, duration-based feature analysis is used, which reduces the number of data points available to train and test models of agitated behavior.

2.2 Respiration Monitoring for Asthma Care

Asthma is a chronic disease of the respiratory system. Asthma attack or exacerbation is a severe acute condition for asthmatic patients. In US, one in twelve children suffers from asthma, whereas 11.5 million people face asthma attacks every year. The risk of exacerbation for asthmatic patients is often associated with short-term or sudden exposure to air pollutants, such as ozone (O3), even indoor [101-104]. Unfortunately, 50% of US population remain exposed to pollutants above the national safety standards. Hence, personal exposure tracking strategies, based on the nociceptive lung function response to pollutants, is essential for asthmatic patients.

Personal risk tracking strategies require continuous monitoring of respiratory parameters,

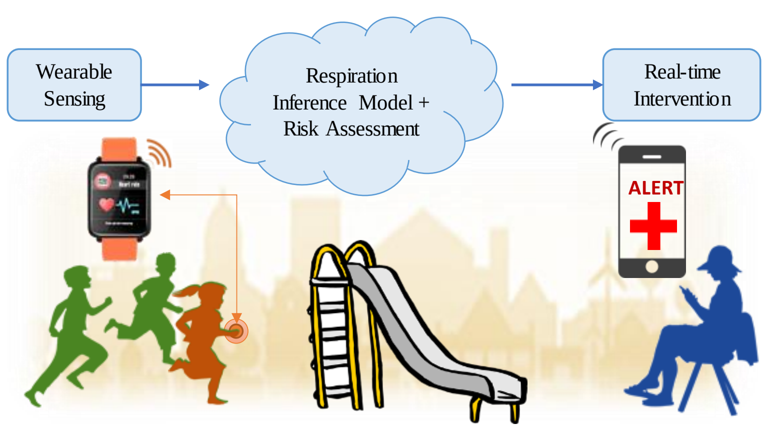


Figure 2.3: The VICTER study aims to infer respiratory parameters from non-respiratory sensing

especially the instantaneous minute ventilation, V_E, which is the amount of air breathed in or out per minute. V_E is a major factor in determining the “effective” dose of exposure; exposure to even moderate pollutant concentration level at high ventilation rate can induce complications in the lung function [105-106]. The effective risk of exposure is directly proportional to the product of the pollutant concentration in the air ($C[O_3]$) and the minute ventilation (V_E). Hence, continuous V_E monitoring can enable potential risk assessment to prevent exacerbation. The asthma care project, named ventilation inference for continuous tracking of exacerbation risk (VICTER), aims to infer the minute ventilation toward assessing the instantaneous risk of pollutant exposure (Fig 2.3). VICTER aims to explore and quantify the risk of ozone-induced environmental asthma, and is supported by the National Institute of Environmental Health Sciences. This project pursues multiple experiments in parallel including the lung function response to ozone exposure and prospective interventions, and the personal and contextual variation in lung function.

Using custom designed wearable devices, the environmental pollutant concentration ($C[O_3]$) can be continuously measured. Commercial ozone sensor MiCS-2614 is used in our wearable ozone monitoring platform. MiCS-2614 was chosen for wearable ozone monitoring since its miniature dimension, low power consumption and high sensitivity. The size of the MiCS-2614 is small with area of 35mm^2 and height of 1.55mm. Inside the metal capillary barrier, there is a sensing resistor R_S to react with ambient ozone, and a measurement circuit to convert the resistance of R_S to an output voltage. During normal operation, R_S will be heated up to 430℃ with typical 80mW heater power. Then ambient ozone diffuses into the sensor through the metal capillary barrier to react with R_S. The output voltage reflects the resistance of R_S which could be employed to calculate ozone concentration level. Ozone detection ranges from 10ppb to 1000pb,

Figure 2.4: Custom designed wrist-worn device (left, center) to monitor pollutant concentration and sense wrist motion, Shimmer ECG sensing unit (right) is worn on chest to sense ambulatory ECG

which is capable of monitoring indoor and outdoor environment. Shimmer 3 is used as a data logger, and a daughter board with sensors and related circuits are designed to connect to a Shimmer 3 node as shown in Fig. 2.4. Three 12-bit analog-to-digital converter (ADC) channels on Shimmer 3 are used to sample sensor data of ozone, temperature, and humidity. The temperature and humidity values are used to instantaneously calibrate the ozone readings.

Spirometry is the clinically accepted standard for measuring respiratory parameters [107]. This modality, even in its portable form [108], is extremely invasive and not suitable for continuous day-to-day use. Hence, out-of-hospital or at-home continuous respiration monitoring remains an open challenge. Along with the direct measurement methods, indirect, or surrogate, measurements from other physiological signals, such as ECG-derived respiration (EDR), are gaining momentum. With the advent of wearable sensors, such methods can achieve the long-sought unobtrusiveness and usability. Yet, the challenge remains to improve the measurement performance against the noise and uncertainty in signals acquired using wearables.

Continuous respiration monitoring has been an active area of research for the last two decades yielding many disruptive technologies [109-112]. Recent research efforts are looking for noninvasiveness and day-to-day usability, either by designing body-worn, contactless sensing devices for direct measurement, or by estimating respiration from wearable sensing-driven non-respiratory signals, to achieve real-world applicability. Direct respiration sensing methods try to capture any related physiological phenomena. For example, inductance plethysmography can track changes in thoraco-abdominal surface area during respiration using two transducer sinusoidal coils and an oscillator on the body [113, 114]. Similarly, magnetometer plethysmography tracks changes in body volume by magnetometer transmitter-receivers. Also, piezoresistive, piezoelectric, and capacitive sensors are explored to capture the respiration-time transthoracic modulation [115-117]. Non-contact modalities such as radar, optical, and thermal imaging have also been proposed to achieve contact-free respiration monitoring. While such methods bring in the capabilities to measure respiratory parameters beyond breathing rate (BR), their performance and usability need evaluation beyond stationary, calibrated, location-specific lab settings across contextual and inter-person variability in free-living.

Recently, research efforts toward estimating respiration as surrogate or indirect measures from peripheral physiological sensors are gaining momentum, thanks to both the technological advances and the wide acceptances of wearable sensors. Modalities such as electrocardiogram (ECG) and photoplethysmogram (PPG) are at the center of these efforts trying to capture the physiological interaction between cardiac and respiratory functionalities. ECG derived respiration (EDR) methods often use signal processing techniques, such as power spectrum analysis, wavelet transform, empirical mode decomposition, to demodulate or extract the respiration signal, and then estimate the related parameters [118-124]. Such methods are often prone to propagation of reconstruction error, worsening estimation performance. These methods can provide some level of interpretability compared to other machine learning methods. ECG features are used with variants of principal component analysis in data-driven models to estimate respiratory parameters [125,126]. PPG-based methods follow similar processing and modeling techniques, while adding the benefit of non-invasiveness by acquiring the signal using wearables [127-129]. Most of these works focus on BR as a coupled parameter of heart rate, which often lack to overcome the contextual variations. Other respiratory parameters, such as tidal volume and minute ventilation (V_E), remain less investigated. A few works on V_E estimation use multimodal sensing with ECG, PPG, and motion from inertial measurement units (IMU), where IMU signals are often used in denoising motion artefacts or extracting activity intensity [130-132]. However, the robustness of such methods against ambulatory noise from wearable modality, and against interpersonal and cross-context variations, is yet to be evaluated in real-world setting.

With the motivation toward asthma attack prevention in the VICTER study, I attempt to estimate

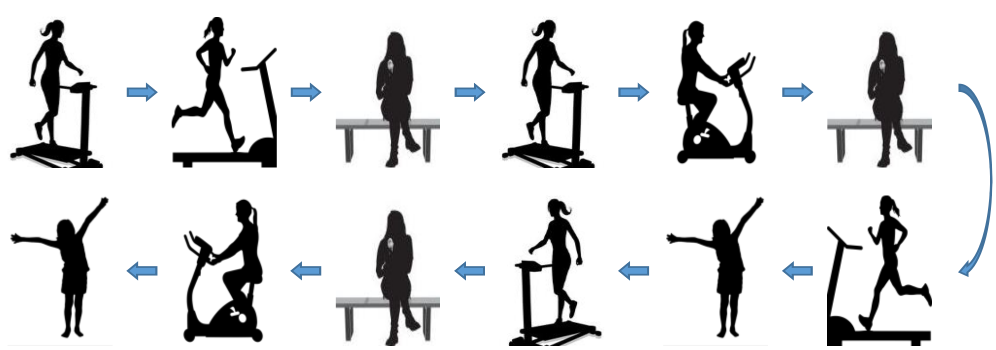


Figure 2.5: Healthy subjects are recruited to acquire respiration data across an exercise protocol

the respiratory minute ventilation (V_E) using wearable ECG and wrist-worn IMU sensors. Challenges toward this objective span from sensor noise reduction to physiological signal representation and modeling the relation-ship between sensor data and respiration. The inter-personal and contextual variations among the physiological parameters challenge the exploration of interpretable relationships. To explore the potential of wearable sensing, namely ECG and IMU, in tracking respiration continuously, I collect wearable sensor and respiration data using a physical exercise protocol (Fig 2.5).

15 healthy volunteers, 9 women and 6 men, participates in this study. The participants come from various ethnicities, and their health and fitness statuses are different. They may have mild asthmatic history, as this is not an exclusion criterion. The only exclusion criteria are pregnancy and/or tobacco use. Table 2.2 presents the demographic details of this population.

Each participant wears two commercially available devices: a Shimmer3 ECG device on the chest and a Shimmer3 IMU device on the wrist (Fig 2.4). The ECG unit is programmed to collect three bipolar ECG channels (Leads-I, II, III). The IMU houses 3-axes accelerometer and 3-axes gyroscope sensors to capture the wrist motion. Both devices have on-board MSP430 micro-controllers that sample the ECG and the IMU signals. The ECG signals are sampled at 250 Hz with ADC gain adjusted to capture 800 mV differential range, and are stored on an on-board flash memory. IMU signals are also sampled at 250 Hz, and are recorded to on-board flash memory.

To acquire respiration measurements as ground truth during the data collection sessions, clinical Spirometers, comprising pneumotachometer (Hans Rudolph model #3830), amplifier (HR PA-1 series-1110), connector (series-7001), 2-way non-rebreathing Y-valve (series-2730), and data acquisition device, are used under human expert supervision. This device acquires BR in breaths per minute, inspire duration

TABLE 2.2 DEMOGRAPHIC COMPOSITION OF THE VICTER STUDY

Sex	Age (year)	Height (cm)	Weight (kg)	BMI (kg/m^2)	Race (White/All)
M (6)	25 ± 5.5	173 ± 4.8	77 ± 16.3	26 ± 5.8	2/6
F (9)	22 ± 3.0	164 ± 7.7	67 ± 7.1	25 ± 4.3	5/9

in seconds, tidal volume in liters, peak inspiratory flow in liters per seconds, and V_E in liters per minutes.

To capture the contextual variations of the lung functions, a multipart physical exercise protocol is designed. Each participant is assisted by an observer in following the protocol step-by-step. Before the experiment, the participant is instrumented with the wearable devices. The protocol uses a treadmill to facilitate some of the activities. The designed sequence consists of 3 walking and 2 running sessions on the treadmill, 2 stationary biking, 2 random hand-waving movements, and 3 rest periods (Fig 2.4). The protocol allocates about three minutes for each activity, as well as a padding of two minutes of rest between consecutive activities. To allow the physiological changes related to an activity reach stable states, we do not collect respiration labels for that the first minute. After performing each activity for one minute, the participant is instrumented with the Spirometer mouthpiece and the nose clip. The participant resumes that activity for about two more minutes during which both respiration and wearable sensor data are acquired. Thus, each data collection session with all the five activities takes about 80 minutes, though both sensor data and respiration labels are acquired for about 20 minutes. This protocol is approved by the IRB of the University of North Carolina (UNC) at Chapel Hill. The sessions are conducted in a specialized physiology monitoring facility at the EPA Human Studies Facility in Chapel Hill, NC in partnership with the UNC Center for Environmental Medicine with a cooperative agreement (US EPA CR 83578501).

Using these two health monitoring applications as context, I pursue the research questions presented in the previous chapter toward real-world long-term sensing data acquisition and utilization for modeling behavioral and physiological parameters.

RELATED PUBLICATIONS:

- N Homdee, R Alam, J Hayes, J Park, S Wolfe, J Lach, "Agitation monitoring and prevention system for dementia caregiver empowerment," Computer, IEEE Computer Society, 2019.
- A Bankole, M Anderson, N Homdee, R Alam, N Fyffe, H Goins, T Newbold, T Smith-Jackson, and J Lach, "BESI: Behavioral & environmental sensing & intervention for dementia caregiver empowerment - Phases 1 & 2," SAGE American Journal of Alzheimer's Disease & Other Dementias, 2020.
- R. Alam, et al., "Wearable respiration monitoring: Interpretable inference with context and sensor biomarkers", IEEE Journal of Biomedical & Health Informatics, Under review, 2020.

Chapter 3

Robust In-Home Sensing

The first challenge I address in this research is the one with data acquisition using wearable sensor devices in real-world settings, more precisely, at unconstrained residential spaces, for long-term deployments. Existing methods on wearable sensing that use commercial platforms such as smartwatches often require carrying a smartphone, so that the watch can transmit data using Bluetooth to the phone for real-time analysis. Another approach is to transmit using Wi-Fi to a server for data processing and model running. The third approach is to process sensor data and to run associated models on-board the watch. While each mode has its benefits and drawbacks, considerations on device battery life and usability burdens are critical for real-world long-term use. Carrying a phone all the time poses usability burden for patients especially those who are elderly and children. On the other hand, communication over Wi-Fi consumes much higher battery power than over Bluetooth. Moreover, Bluetooth streaming requires less compute and memory resources on the watch [136]. These usability and resource constraints impact the system performance in the real-world settings and affect the data acquisition pipeline. Many existing out-of-lab applications use the watch-with-phone mode of sensing systems. Such applications impose some usability constraints such as keeping the phone nearby, keeping both the watch and the phone charged, memory and compute resources on the phones, and so on [74]. For long-term deployments, such burden may affect user participation and data quality and quantity.

Another challenge for long-term data acquisition is maintaining the system operation for the duration of the deployments. Once a system is deployed, to ensure reliable data collection, some mechanism needs to keep monitoring the deployed monitoring system. Hardware faults, software crashes, network disconnections, human interference, and many other reasons can result in the deployed system being partly or completely inoperative with accompanying loss of valuable data. In many of these systems, once the system is deployed in a home, it is either not monitored, or a system administrator manually remote monitors the application monitoring system. For real-world long-term deployments, dependence on manual monitoring may risk the system operation.

As the first research question requires, under the usability and resource constraints, I identify the sources of data loss in residential deployments of sensor systems. And I propose robustness and usability measures to reduce some of those losses. In this endeavor, I use the BESI study on dementia care as the real-world setting and design the BESI system to experiment on these challenges and to demonstrate the effects of the proposed solutions.

3.1 Failure Modes in Residential Deployments

Residential deployments of wearable sensor systems face various challenging scenarios, caused by both the system and its surrounding environment, including the users. These challenging scenarios may introduce failures and impact various parts of the system and its environment, which may cause loss of data, loss of observability, improper, even safety critical, actions may be taken in case of an intervening system [17-19]. All of these failures may also not be immediately observable depending on the monitoring strategies employed, some may surface in the data or surveys at post-deployment analysis. From the BESI deployments, I discovered some major failures that needs to be addressed by the systems design and architecture. Those failures are presented here based on the associated components of the system.

3.1.1 Wearable and Mobile Related Issues

Since various motion and physiological sensing modalities are incorporated in commercially available smart watches, phones, fitness trackers, etc., more and more research efforts are employing these devices instead of custom-built platforms. Such practices provide some level of ease and reliability, reduce time and effort spent in infrastructure development and allow more for advancing science and applications. But these also bring in various limitations and restrictions such as app permissions, sensor configurations, networking and connectivity, power usage, priorities, etc., in other words reduced observability and controllability. These characteristics impact these devices during deployment in various forms.

a. *Permissions and Updates:* App permissions are used in third-party developed apps for accessing APIs related to sensors, Wi-Fi, Bluetooth, etc. These permissions are maintained and controlled by the owner company. Because of privacy and security concerns, some app permissions used to be allowed may be deprecated. Such sudden policy changes impact a deployed app. Similarly, for pushed OS or system software updates, some apps may stop functioning as designed. In one of the early deployments of the BESI system, updates for the Pebble app on the tablet was accidentally installed by the user, which caused the tablet to forget the paired watches, and required re-authentication and reinstalling the watch apps.

b. *System Settings*: Battery life remains a major concern for wearable and mobile devices, both to researchers and to users. Power optimization algorithms are incorporated in apps and services to ensure longer life cycle of sensing and operation. While deployed, users may inadvertently turn on power saving mode of device. Also, pushed updates of OS or system software may reset the device settings and turn on such features. As a consequence of such incidences, deployed third party apps maybe turned inactive and prevented from collecting sensor data or connecting to the network. In BESI deployments, two Android apps on the tablet and two watch apps (for patient and caregiver) were used. In multiple instances, the survey app automatically closed because of user turning on the power saving mode.

c. *Connectivity*: Connection to Wi-Fi or Bluetooth devices are maintained in the lower level by the OS. Inherent battery optimization mechanisms turn these connections down during hours of inactivity. Also, network connection and security are handled by the system software and are not easily accessible by developed apps. These causes challenges for ensuring continuous observability and data transmission for devices. BESI deployments have experienced the resetting of the BL connection between the relays and the watches multiple times. Likewise, sometimes the watch was stuck in a state where it was unable to collect and send data even after being in range.

3.1.2 Challenges with Embedded Platforms and Servers

Embedded platforms are mostly used to host commercial off-the-shelf sensors and/or low-complexity computing and networking tasks. Such platforms often use open source operating systems and software, hence are less prone to the type of failures presented for wearable and mobile devices. But such platforms pose a different set of challenges involving the methods employed for pipelining computations and the processing complexity and resources. Servers and base-station devices need to address some similar challenges, but given they are equipped with more resources, these challenges are less common on servers. These characteristics impact these devices during a deployment.

a. *Processing Pipelines and Threading*: To incorporate parallel and independent sensing and processing pipelines on the same platform, each pipeline is implemented as a separate thread. But, in case of a failure in any part of a thread, all components of that thread fail, even other threads on the platform are affected indirectly. For example, each of the relay nodes in the BESI system supports heterogeneous sensing using its ADC, GPIO, and I^2C peripherals. While each of these peripherals are sampled on different threads, in case of the ADC failure for the Temperature sensor, the sampling for Audio and Doorway sensors are affected. Also, the "heartbeat" packets from that thread to the base-station decouple from time synchronization process and participate in error message transmission. This communication increases the delay for other threads and leads to missing samples.

b. *Resource Limitation*: Embedded platforms are equipped with lower computational power and memory, which makes them more prone to failures associated with processes that tend to utilize the bus speed and memory extensively. On the BESI relay nodes, parallel threads are implemented to sample the sensors and to process the data for feature extraction, especially the Audio sensor data. This process is critical not only for data collection, but also for preserving privacy by extracting some essential features from the data and deleting the raw data from the relay node. Even though the computational complexity is well within the resource limitation of the relay nodes, during deployments we have experienced delays caused by accessing memory units to fetch and delete data at a higher speed. This issue causes a delay for the thread to hand-over and consequently affects other threads and causes data loss.

c. *Interoperability and Compatibility*: While open source resources and packages are very useful in quicker implementation of programs and prototypes, the reliability, efficiency, even computational requirements are not always properly mentioned to use for optimal usage. For example, feature extraction using open source signal processing packages on the embedded platform sometimes outputs different results that that generated by MATLAB signal processing library on a server.

3.1.3 Network and Connectivity Issues

Network is a vital element for residential sensing systems. The challenges associated with the network has impacts on multiple devices connected to the network. Network and connectivity failures not only cause data loss but also impacts observability of the subsystems.

a. *Scaling*: Network is a shared resource. Dual-band routers, after configuration, uses a single channel to establish the network between all devices. This fact poses a limit on the scaling factor for a residential sensing system. The failures associated with this challenge are very disperse in nature, as one component adding overheads on the network affects the network quality for other devices. In BESI deployments, this issue kept ramping up with the size and physical coverage area of the system. Often a relay node would not be able to connect to the base-station when other relays are continuously using the network. This failure is also Byzantine in nature, as the relay may show up as connected on the router, but is not virtually accessible though the network.

b. *Adapter and Driver Quality*: While commercial products provide relatively reliable network adapters and drivers as well as less controllability on these low-level components, customized embedded platforms need the researchers to ensure those features. As failures associated with these components often require solution such as hard reset, which causes the whole device to stop functioning for the time being. In BESI, we have experienced this issue not only with our Wi-Fi adapters, but also the Bluetooth devices. And, resetting the adapter requires re-authentication which is harder to handle remotely and automatically.

c. *Induced Time Synchronization*: Residential sensing systems employ many devices and platforms, many of those do not have a real-time clock (RTC) on-board. Hence, network delays and packet losses may impact the time synchronization process induced from a central server with an RTC. The severity of lag in this process depends on application.

3.1.4 Remote Monitoring Challenges

Deployment restricts the ability to physically access the system for any kind of diagnosis and recovery, making remote monitoring an essential component in practice. The goals of remote monitoring are to ensure that the system is operating as expected and to detect any failure as soon as it occurs. The challenges associated with this depends on the tools employed, and these failures impact the system from observability to recoverability. While an ideal scenario would be to have a fully automated monitoring system, experiences with “false alarms” and “hidden failures” reduce the credibility of such automation.

a. *Byzantines*: Sensing happens on different components of the system, while monitoring is performed at a higher-level dashboard on a server or the cloud. Consequently, failures may remain hidden using the favor of abstraction by the dashboard program. In BESI deployments, one such failure was associated with individual Temperature sensors recording errand values due to physical dislocation. The ADC thread on the relay recorded those errand values and sent a “heartbeat” message to the base-station dashboard program. The abstracted relay status (operating or not) on the dashboard causes this fault go unnoticed.

b. *Context of Failure*: Monitoring logic needs to incorporate the context of the data into consideration. For example, the dashboard sends out an alert when the Pebble is disconnected from the relays. But this may have caused because the user leaving the premise wearing the watch, which is not a scenario to be alerted. This “false alarm” can be avoided by incorporating context to the failure model.

c. *Network Overload*: To achieve ease in monitoring, debugging, and recovery, it seems lucrative to request maximum observability of the system, for example using remote visual access to the local server through remote desktop software such as TeamViewer. But such media impact the network adversely by adding video traffics to the internet. Also, in case of user paid Internet connection, this may reduce participation enthusiasm.

3.1.5 User Participation Difficulties

A major obstacle in real-world deployment scenarios is user participation. The users are often requested to provide ground truth information to validate claims made from the collected data. A low user compliance could affect the sensor system ability to provide meaningful labels to the collected data which makes the sensor data collection ineffective.

a. *Ground Truth*: To provide the collected sensor data with meaningful ground truth labels, the BESI study requests the dementia caregiver to report the time of patient agitations. These agitation reports are done via survey through tablet application. Typically, the agitation reports are done after the dementia agitation occurred and dealt with. This causes a slight shift between the time that agitation is reported and the actual time that the agitation has happened. This is due to human error, since the caregiver could never remember the exact time of agitation event. The time shift in the report can mislead the BESI study into observing the environmental context at the inaccurate time.

b. *Aesthetics and Unobtrusiveness*: The appearance of the devices deployed at the residential site is a challenge towards the user's willingness to participate. The sensors deployed in residential environment should be unobtrusive to the daily lives of the user. In BESI, one of the caregivers reported that

… after a few weeks, the patient with dementia was asking when would the sensors be removed.

This highlights the challenge due to the patient noticing the sensor devices and wanted them removed. In another deployment, some sensor devices are reported to be obtrusive due to the users came in contact with it and causes the device to be knocked out of its placement,

… sensors were in travel paths and were pulled off wall.

c. *Training and Usability*: Most of the environmental context in the residential system can be passively sensed. However, some essential information could not be collected via sensors. An example of this information is an agitation event report from caregivers. As mentioned in the previous section, these event reports function as labels for the collected data streams. In BESI, a caregiver wearable is provided for marking the time of agitation instantly. Caregivers are also requested to report agitation events through the provided tablet application for more detail of the event. This required caregivers to be familiar with the provided technology. This challenge hinders a successful deployment if the user could not use the technology and provide meaningful data labels.

d. *User Compliance*: The deployments often require certain assistance from the users in order to succeed. In BESI, the inertial motion of the patient with dementia is collected via a smartwatch. Due to the continuous streaming of the motion data, these wearables have a battery life time of 16-18 hours per charge. This requires a compliance from patients to wear the device throughout the deployment and swap them once the battery is low. Assistance from caregivers was also needed to keep these wearables charged. This challenge was emphasized by one of the caregivers on the patient compliance regarding the wearable:

... patient did not wear it all the time because it was not her regular watch.

3.1.6 Challenges with Deployment and Operation Management

For successful data collection over long-term deployments, the deployment process and the operation management need to be considered. An unprepared deployment could result in the installation crews spending too much time at the residential site, causing burdens to the users. In addition, a monitoring procedure is also needed after the deployment in order to ensure sound operation of the system.

a. *Deployment Preparation*: Real-world residential sensing systems are often aimed to be deployed for several weeks. This long period deployment is a challenge due to the long operating time of the sensor devices. In BESI, long period test deployments were done in-lab before any actual deployments. This makes sure that the sensors devices are ready to be deployed and can operate for an extended time. It also ensures that the equipment needed for an actual system are available. In case of unexpected challenges during an actual deployment, extra sensor devices, electrical cords, outlets, etc. are prepared.

b. *Deployment Time and Labor*: In the process of installing sensor devices at residential sites, the deployment time has to be considered and used effectively. An inefficient use of time during deployment could disrupt the user's daily schedule. Typically, two to four team members are responsible for each BESI deployment. Half of the crew focuses on installing the sensor devices throughout the residential site, while the other half are responsible for testing the deployed sensing system. Typical challenges during an actual deployment are: unobtrusive placement of relay stations, insufficient power outlets for components, and the

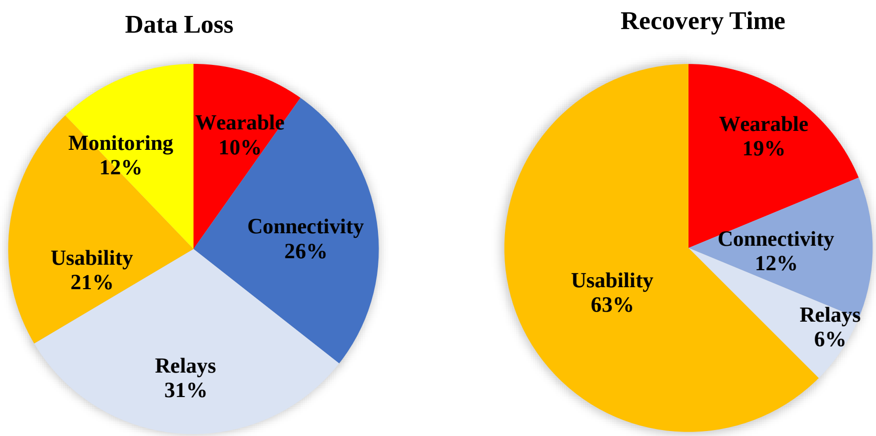


Figure 3.1: From the BESI deployments, data loss and recovery time associated with each failure mode.

diversity of the size and shape of residential settings. Some unexpected challenges could occur at the household such as flighty internet access or faulty behavior in the sensors. In BESI, the average system installation and testing was around 1.5 hours.

These failure modes answer the research question regarding the source of potential challenges in real-world residential deployments (Fig 3.1). As a solution to the above-mentioned failure modes, I propose and iteratively design the BESI system to facilitate real-world long-term continuous deployments and enable data collection from wearable sensors in residential settings.

3.2 The BESI System

The BESI system is proposed as a novel wearable-edge interconnected system for home setting. The IoT revolution has initiated an enormous emergence of miniature computing devices, from smart thermostats, smart meters, to smart TVs and garage doors. BESI aims to leverage both the computing power and the wide-spread placement across residences of such devices for overcoming the limitations of power, computing, and connectivity resources of smart watches, as well as for implementing a distributed autonomous system maintenance framework. The BESI system components and their functionality toward reducing the real-world failure modes are presented in this section.

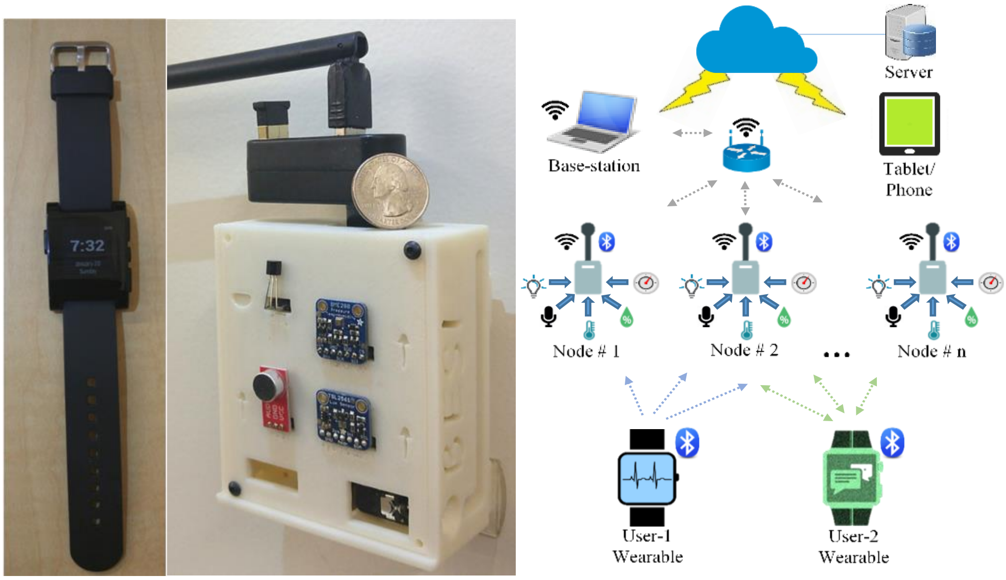


Figure 3.2: The BESI system: (left) Pebble watch and the custom edge node, (right) the network architecture.

3.2.1 System Components

BESI comprises of wearable devices, room-level edge nodes, a local (base-station) and a cloud server, and interactive modules (tablet or phone). Fig 3.2 shows the system architecture and the interconnection of the components. BESI utilizes multimodal local wireless networks and an Internet connection to establish a cyber-physical IoT platform. To add reliable system operation for long-term, it adapts the Monitor[2] platform to achieve hierarchical monitoring and autonomous recovery [85,86]. The components of the BESI system are described here with their functionalities.

3.2.1.1 Wearable Sensing Devices

A commercially available smart watch, made by Pebble Technology Corp., is used to sense the inertial motion of the user from their wrist. The Pebble watch features an ARM Cortex-M3 processor and a 3-axis accelerometer with programmable sampling rates of 25, 50, or 100 Hz. It uses Bluetooth 4.0 (BT) for communication.

Sensing: To collect the wrist motion data of the user, I developed a background service named Pixie. This service samples the accelerometer at 50 Hz and sends that data to any connected edge nodes in batches at 10 second intervals using the data logging API. Pixie incorporates a companion foreground app that displays a standard watch-face and handles accidental button presses.

Connectivity: To maintain connectivity with the edge devices, the background service works as the manager (master) of the BT connections. The manager closes an existing connection at any point of packet transmission failure for 2 minutes. After connection loss, the service scans for authorized edge devices. Based on the received signal strength of the broadcasts from the edge devices, the watch attempts to establish connection with the nearest available edge device. Out-of-home scenarios are detected from the scan results and handled by first stopping the sensing component and then continuing to scan at a lower duty cycle every 3 minutes.

Maintenance: The Pebble watch is programmed with both self-monitoring and hierarchical message passing features. A companion service ensures that the Pixie service is always running in the background and restarts it at any point of crash or power reboot. It can reboot the watch if Pixie keeps crashing. The exception handling on Pixie is designed in a step-wise manner to handle the sensing and connectivity failures. Restarting the service addresses the sensing software failures, but for hardware fault, the service will keep closing, and notify the edge nodes. The connectivity failures and handover attempts are managed by the Pixie service itself. It will close the service in case of multiple failed connection attempts or exceptions from the BT connectivity driver.

3.2.1.2 Edge Computing Devices

BESI uses a network of edge computing nodes distributed across the residence of the users (Fig 3.3). These nodes enable continuous connectivity of the wearable devices with the server and reduce the sensor data processing load from the wearables. Each edge node is implemented on an embedded computer board, the Beaglebone Black. The board features an ARM Cortex-A8 processor, 512 MB of RAM, and 4GB of flash memory. It provides interfaces like GPIO, ADC, SPI, and I2C for peripheral devices and is capable of incorporating both BT and Wi-Fi connectivity. The processor runs Debian, a Linux-based operating system. Each node is equipped with a 32 GB SD card to facilitate data collection for long periods. Each edge node performs the following tasks:

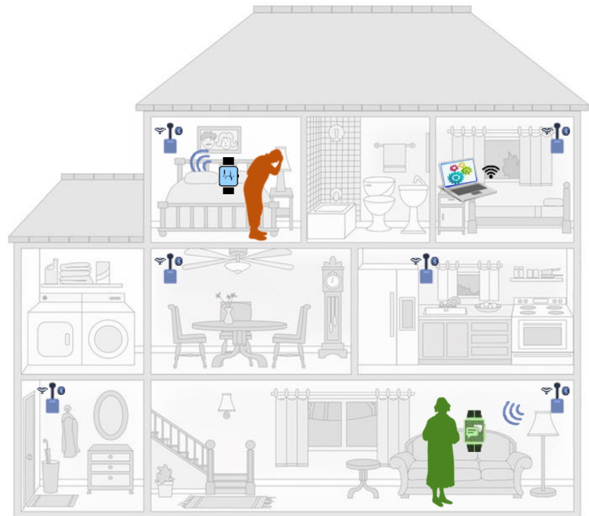


Figure 3.3: The BESI system is spread all over the residential space to facilitate seamless health monitoring.

Connection to Watch: The edge nodes connect with the Pebble watch using BT. Each node hosts two parallel processes, one to maintain continuous connection and the other to maintain the data link to the watch. The connectivity process broadcasts connection request if no watch is connected, and handles authentication and establishes data link when a connection is requested by the watch. The data logging process uses an open source python library Libpebble2 [137], which provides tools to implement several Pebble protocol services including the data logging protocol. This protocol is modified to acquire sampled accelerometer data from the watch. I also incorporated the ZeroMQ library [138] for handling data reception in batches and message passing to and from the watch. These developed processes run in parallel on the nodes and store data on the SD card.

Edge Computing: The edge nodes are programmed with two parallel processes to perform additional data processing tasks. One process handles the collected sensor data by filtering, denoising, and time synchronizing the wrist motion signal. The other process adds feature extraction functionalities on the motion data to reduce the network load of forwarding raw signal streams to the server. This process reduces the memory and computing restrictions from the watch, as well as the workload and latency of the data modeling pipeline on the server. The edge computing platform supports additional processes to be easily implemented without affecting the performance of existing processes.

Connection to Server: Each edge node connects with the server over Wi-Fi. This connection is used to transmit the extracted features to the server as well as to communicate with maintenance packets. Each node sends a low-frequency maintenance packet to the server stating the status of the running processes, connectivity status to watches, and some statistics about the data quality at a lower frequency like a 'Heartbeat' message.

Maintenance: The edge nodes locate in the middle of the connection between the watch and the server. Hence, these need to cover major system operation maintenance responsibility. Independent processes are designed to monitor the watch-edge and the edge-server connectivity and data link. For the watch-edge connection process, failure to connect to consecutive requests from the watch generates an exception. Similarly, failure to connect to Wi-Fi while there is a network available causes an exception. Such exceptions make the respective processes to restart their connection modules. Continuation of these exceptions cause the process to close. Similar watchdog processes are implemented for the data reception (from watch) and transmission (to server) pipelines. A dedicated process looks at the logs of these processes to find out continuation of process crashing. In such scenarios, the device is reboot with a flag to the maintenance module on the server. Finally, a process control system named Supervisor [139,140] manages the running processes on each node. It is configured to monitor and manage all the connectivity, logging, and maintenance processes. It ensures that all processes are running in their regular courses, restarts in case of any process crash, and starts all processes after system reboot. New processes can be easily added to the Supervisor watch-list.

3.2.1.3 Local Server

The local server is implemented on a laptop with an Intel i7 2.8 GHz processor, 12 GB of RAM, 300 GB HDD, and Windows 10 as operating system. It is equipped with internet connectivity to send out system maintenance alerts to the human operators as well as to send out model-driven notifications to the phones or tablets of users. The local server performs the following tasks:

Inference and Intervention: The inference engine is implemented on the local server. After the training phase, the model is packaged as an executable file and is called from a process running on the server. The engine predicts about human attributes based on the features received from the edge devices.

The prediction parameters are passed to an intervention management process for generating notification. The intervention manager incorporates heuristic and contextual information with the prediction outcomes to decide about notification selection. The notification is sent both to the smart watch via the edge devices and to the tablet or phone via internet using the Firebase Cloud Messaging API.

Connectivity: The server maintains parallel communication channels over Wi-Fi with the edge nodes using the python socket library in a dedicated connection manager process. These connections are used for monitoring and managing the operation of the edge nodes and the watches, for receiving the features extracted from the watch motion signals, and for transmitting the intervention notifications to the watches. The connection manager is also responsible for synchronizing the motion features received from the edge nodes. The internet connectivity facilitates remote monitoring functionalities and the intervention delivery to the interactive modules like the tablet.

Maintenance: The server leads the over-arching maintenance functionality of the whole system. The monitoring process autonomously maintains the connectivity modules and the operation of the inference engine and the intervention manager. In case of any connectivity failure, the monitoring process can switch, disable, and restart the connection ports. It sequentially reboots the router and, if needed, the server, for sustained internet connectivity issues. All the processes are maintained by a start-up process manager for ensuring continued operation. In case of any data inconsistency, the monitoring process attempts to redirect the watch connectivity to a different edge node. Similarly, the monitoring process on the server hierarchically manage the rest of the systems operation using the ‘Heartbeat’ status updates from all the edge nodes. In case of sustained issues, the monitoring process sends an email to the human operator with related status codes. A remote access control host service, TeamViewer Host, keeps running on the base-station. This service allows the human operator secured remote access to the base-station from any device with a client service and authentication information. Edge nodes operation can be debugged by the human operator through remote SSH connection to those devices.

3.2.1.4 Interactive Modules

To enable the users to provide labels about any related events, BESI incorporates two independent modules: a tablet and a smart watch, both hosting application-specific apps. These interactive modules are also required for providing interventions to the user in an effective, timely, and convenient manner.

Smart watch: An app called Memento runs on the Pebble watch of the user. This app is designed to collect label or timestamps for an event. To mark the time of an event when occurring, the user presses any button and the Memento app registers that information on-watch and sends that to the tablet app via the edge devices and the servers. The watch app also receives intervention from the servers via the edge devices in the form of both haptic and text notifications.

Tablet: The tablet hosts an Android app that allows users to input details about events more than just label timestamps, for example qualitative information about an event or usability feedback about the system. The app contains a brief survey designed to collect that information. The tablet app also hosts the interventions and notifications sent from the servers.

3.2.2 Robustness Features of BESI

The BESI platform is developed to address challenges in residential deployments of health monitoring systems, like reliability, unobtrusiveness, heterogeneity, and maintenance. BESI incorporates the Monitor2 mechanism to facilitate reliable monitoring and notification. These considerations have influenced the architecture and design decisions of the system, and the strategies for overcoming these issues are implemented as advanced features of BESI (Fig 3.4). The deployments of the BESI system has allowed us not only to experience the failures of various kinds, but also to evaluate various strategies to solve those issues both during run-time and post-deployment design iterations.

3.2.2.1 Strategies for Wearable Devices

Watches and tablets mostly depend on the designed apps that run on those devices. Hence, we focused on incorporating features in the apps to address the challenging scenarios.

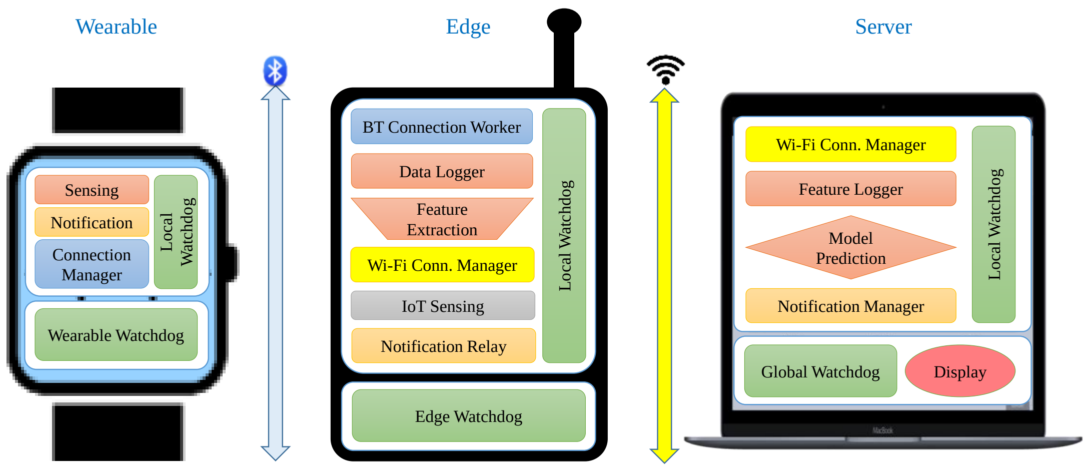


Figure 3.4: The functionality and reliability operators of each component of the BESI system

- *Proactive Manager*: To address the failures associated with connection issues of wearable and mobile devices, we designed and implemented connection managers for both the smartwatch and tablet. The relays host the watch connection manager which attempts to maintain continuous connection. The tablet connection manager is part of the Android app and keeps it connected to the subnet. Also, we designed the apps with an associated background service. The service keeps watchdogs for any accidental app closing and restarts the app if needed.

- *Utility versus Blocking*: To avoid the unexpected scenarios associated with software updates, we have the automatic updates blocked. Considering the long data time for watch reset, we also disconnect the watch from the tablet to avoid watch reset. In addition, to avoid users accidental tempering of the watch or tablet and closing the apps, we reduce the utility of the devices by blocking other usage of the devices and forcing the devices only to be used as intended. While this enforcement reduces the utility of the devices to the user, it also prevents failure scenarios and helps with the study.

- *Mobility with handover*: To maintain continuous connection between the watch and the relay nodes when the user moves around the residence, most systems impose unrealistic restrictions on range or distance from a fixed location receiver or require the user to carry a phone or other receiver. BESI relay nodes are furnished with processes to establish a direct connection to the Pebble using the serial transport channel implemented by Libpebble2. This transport uses the built-in BT serial support to communicate with the watch. All relays attempt to establish a connection with the watch continuously, but Pebble allows only one connection at a time, and enables its BT search and advertisement only when it remains disconnected. Thus, when it moves out of range from a previously connected relay, a new relay in range establishes connection and starts collecting transmitted data. Pixie is designed to resume sensing and transmission on connection.

- *Scalability and Generalizability*: Each relay possesses multiple peripheral ports to accommodate in parallel many sensing modalities both digital and analog, thus the sensing functionalities are easily scalable and generalizable for various health applications. The number of devices that can connect to the network is not strictly restricted as additional nodes add minimal network traffic burden. Since the watch can switch connection smoothly among relays and each relay cover a long BT range, addition of more relays can cover bigger physical space. Multiple wearables can also connect with the system without any loss of functionality by redistributing connection host assignment. BESI has been used with a Shimmer3 [] as wearable without losing any functionality.

3.2.2.2 Features for Edge Devices and Servers

Relays and base stations are prone to failures associated with program crash, error handling, and resource management.

- *Error Models*: We designed error models to check for errant values from each sensor and incorporated on the sensing relay thread. Such models send "heartbeats" about the error to the base-station and can be detected faster.

- *Multithreading & Parallelization*: Dependable sensing is provided by BESI with implementation of multithreading techniques. On a relay node, environmental sensor signals are sampled, preprocessed, and stored. These data are analyzed on-node in real-time for predicting any health crisis event. All these tasks are implemented as independent parallel processes and separate threads handle different modalities sensing and processing. Hence any intermittent, transient, or permanent hardware fault on a single sensor cannot affect data collection and processing from other modalities.

3.2.2.3 Network and Connectivity Solutions

To address the network channel overloading and scaling issue, we employed two strategies that helped us overcome the failures. We considered the mechanisms behind channel interference and overloading and their effects on channel quality, signal strength, and link loss rate.

- *WAP and Range Extender*: Since the network cannot unlimitedly scale due to physical resource constraints, we incorporated a dual-band extender to increase the coverage area of the network. While the central router or wireless access point (WAP) used the 2.4GHz band channel 11, the extender was configured to use channel 1 of that band to avoid interference. The "back-haul" between the WAP and the extender used the 5GHz band. Using this strategy, we had been able to reduce the relay connectivity issues by almost 80%. While the mobile devices have built-in adaptive network selection strategies, this approach didn't affect the network connectivity of those devices.

- *Subnet Masking and Static IP*: The private IP for most home networks are set as either 192.168.X.X or 10.X.X.X or 172.16.X.X. When deploying our system in a residence with such an existing network set up by an ISP, we experienced that creating a different subnetwork for the BESI system helps to avoid confusion in monitoring and debussing in run-time. We avoided those three private IP subnets and used a different IP for our local subnet. Also, we configured the network with a static IP for our in-situ devices. This configuration accelerated the network setup and evaluation during a deployment visit.

- *Decentralized Network Architecture*: Available systems mostly present central server dependent network architectures, where sensing subsystems transmit data to a central device and that central node processes all streams []. Reliability of such an architecture is dependent on stability of network links and the central node. Failure of the router or the central device may disable the whole system. Residential deployments are highly prone to both accidental device failures and network disconnection []. We designed BESI as a network of independent relay nodes. This architecture is tolerant against network faults and individual device faults. In case of base-station failure, the relays continue to collect and process data from room-level sensors and wearables independently and thus achieve fault containment. Relays can still communicate with each other and propagate an intervention to a connected caregiver watch based on on-node event detection. In case of a network outage, data collection and event detection continue. Only multihop intervention may get affected for such failures, which can also be overcome by adding conditional connectivity settings for wearables.

- *Network Data Load*: Centralized network architectures add heavy traffic burden on the network channels with increases in system size and therefore are not realistically scalable. BESI relays do not transmit raw data to the base-station. Hence the network load is very low, which ensures connectivity and transmission performance of the network.

3.2.2.4 Remote Monitoring Strategies

Monitoring has two major objectives: failure detection and recovery. We found that full automation is not an optimal approach for monitoring in real-world.

- *Hierarchical Monitoring*: To avoid generating "false alarms" while attempting to detect possible failures, we implemented hierarchical monitoring strategy. The lower-level sensor failures are monitored by the higher-level sensing thread, and the "heartbeat" messages are used to propagate that information to the upper layers. Similarly, the network and the embedded platform issues are detected by the base-station, along with the file-system and peripheral issues, and sent uplink. Such hierarchical message propagation adds confidence in error detection and efficiency in recovery methods.

- *Human-in-the-Loop*: Besides an automated solution, we used human experts to monitor the system on schedule to help the system reduce the confusion added by the uncertain scenarios.

- *Reliable Connectivity*: Each relay node continuously monitors its connectivity with the local router and attempts to recover any disconnection by resetting. 'Heartbeat' messages from each relay contain information about connectivity of that relay with the local network. That is used by the base-station to reconnect if required. The base-station is furnished with a redundant Internet connection by a cellular service to ensure reliable connection. This ensures robustness in remote access to the base-station for monitoring and maintenance purposes.

- *Watchdog and Recovery*: The Supervisor system works as a watchdog on each relay node to keep track of all the sensing, connectivity, processing, and storage processes. These processes are implemented to perform real-time stateless data handling. In case of any process failure, the Supervisor restarts that process. Manual recovery procedures on the relays are executed through remote terminal connections using SSH if they are connected to the local network.

3.2.2.5 Approaches toward User Participation

User participation is necessary for success in studies and applications of residential sensor networks. Increased success in the collection of ground truth and user compliance are achieved by incorporating the following strategies:

- *Multimodal Validation and EMA*: To assist users to provide accurate data labels and ground truth, we incorporate multimodal annotation in BESI. While the user can submit post-event surveys to report an event, the user watch app is also designed as a momentary event marker. These two modalities are matched in post-deployment analysis to validate the ground truth events. Another strategy we discovered as effective is by proactively querying about the labels of a possible event observed in the data. Such proactive prompts are momentary and event-triggered, and it boosts the user participation in providing data labels.

- *Routine and Alerts*: Though each Pebble watch lasts about 16 hours, we provided the user with two watches for their ease to follow a routine of swapping the watches twice-a-day. This routine increased user compliance with watch charging and consequently helped data collection. Also, to avoid missing long duration of data because of watch connectivity issues, we proactively alert the users about watch battery status and query about watch context.

- *Appearance of Wearables*: Users are often not willing to use a visible medical device in day-to-day life. To avoid data loss, we designed the system to be interoperable with trendy smartwatches that transparently contain physiological and inertial sensors. Pebble watch is one such platform that can "disappear" as a smart wrist watch. The Pixie app shows a standard watch-face on the foreground and collects inertial data in the background.

- *3D Printing*: To address the issues with aesthetics and failures associated with physical fall-outs, we designed robust and concise 3D casing for our custom doorway sensor and embedded platforms.

- *Longevity*: Battery life and power consumption of smartwatches affect the usability of these devices. Most continuous-sensing smartwatches provide only about 5-6 hours of battery life, due to the

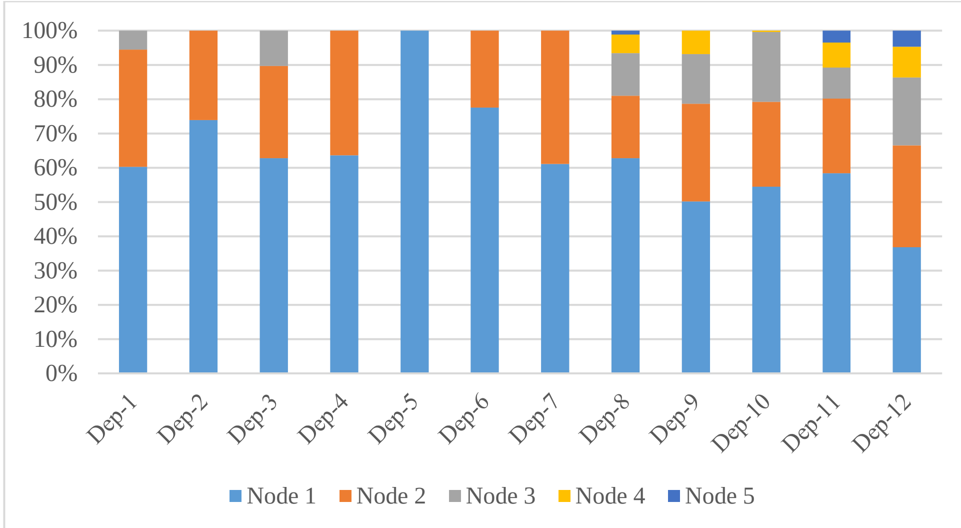


Figure 3.5: Percentage of time the smart watch streamed data to each of the edge nodes

large power consumption by sensors, display, and data streaming. Pebble is equipped with a small 130 mAh Li-ion battery, but it has a low power LCD display. Also, in BESI, the Pixie app batches up collected data for one second before transmitting and thus reduces power consumption. It achieves about 18 hours of battery life between charging sessions of an hour.

Thus, the BESI system is iteratively improved to learn these robustness strategies from the deployments, which are essential for improving system reliability and performance. We conducted such iterations throughout the life-cycle of BESI to establish it as a successful residential sensing system.

3.3 Performance Evaluation

The BESI system is designed to acquire wearable sensor data from long-term real-world deployments in home setting. The system has been deployed at 12 patient residences, the deployment durations ranging from 30 to 60 days each. There are two major aspects of the performance that are demonstrated here: how much wearable data BESII could collect overcoming the connectivity limitations, and how much data loss can it prevent by adapting the autonomous monitoring and recovery strategies.

Due to the variations in the size and floorplans of the deployment settings, different number of edge nodes were deployed (Fig 3.5). During those deployments, the smart watch remained connected to

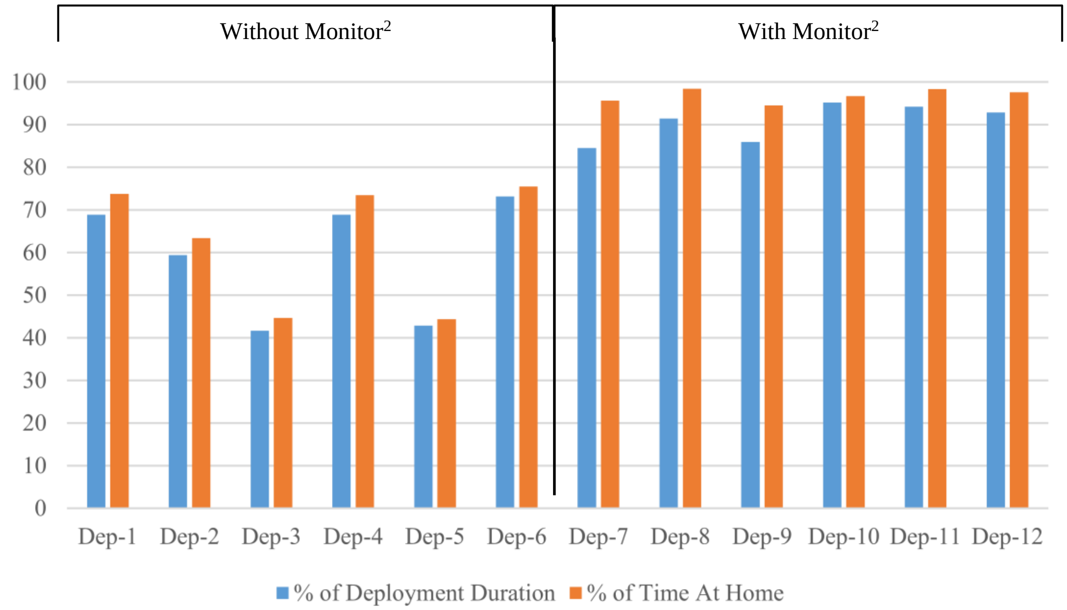


Figure 3.6: Percentage of data acquired over the whole deployments without and with Monitor^2

one node for about 60% of the time, which highlights the challenge mentioned before about the constraint of carrying a phone. Another point to emphasize here is that the percentage of connectivity is not inversely proportional to the number of edge nodes deployed, rather the users' mobility and health behavior play some role in this issue.

The effectiveness of the hierarchical monitoring and autonomous recovery techniques can be demonstrated by the total amount of data acquired over the course of those deployments. To quantify the performance, I propose data availability as a metric for quantifying the robustness of sensor systems in data acquisition. Similar to the definition of system availability, the data availability depends on the duration the system is operating as expected, i.e. the system is available for data acquisition, and the duration the system is expected to operate, i.e. the system is deployed. The definition can be formulated as follows:

$$\text{Data Availability} = \frac{\text{Data Collection Duration}}{\text{Deployment Duration}} = \frac{\text{Data Samples (count)}}{\text{Deployment Duration} \times \text{Sampling Rate}}$$

$$\text{Operational Data Availability} = \frac{\text{Data Collection Duration}}{\text{Operational Condition}} = \frac{\text{Data Samples}}{\text{At Home Duration} \times \text{Sampling Rate}}$$

Using these metrics, the BESI system with the Monitor^2 mechanism has improved the data availability from about 55% (without those robustness mechanisms) to about 95% as shown in Fig 3.6. Without such measures about 45% of the data would be lost due to failures in different parts of the system over the duration of the month-long deployments.

In summary, the BESI system is designed as a novel platform to enable long-term continuous real-world data acquisition from wearable sensors in home setting. This system overcomes the limitation inherent in commercial smartwatch-based solutions and leverage an in-home network of edge devices to maintain constant connectivity with the watch and to alleviate the computing and memory constraints. Moreover, using hierarchical monitoring and autonomous recovery methods, BESI brings ease of maintenance for the researcher team and ensures long-term reliable system operation. With these features, the system has demonstrated success in acquiring more than 90% data from wearable sensors in real-world long-term deployments.

RELATED PUBLICATIONS:

- R Alam, et al., "BESI: Reliable and heterogeneous sensing and intervention for in-home health applications", IEEE/ACM Int Conf on Connected Health: Applications, Systems and Engineering Technologies (CHASE), IEEE, 2017.
- M Ma, R Alam, et al, "M2G: A monitor of monitoring systems with ground truth validation for research in residential applications", IEEE International Conf on Mobile Ad Hoc and Sensor Systems (MASS), IEEE, 2017.

Chapter 4

Sensor Feature Spaces

Wearable sensors capture different aspects of physiology and behavior of human body. To extract data representation from those signals, I employ feature extraction and engineering methods. For the BESI study, I collected 3-axes accelerometer signals from the wrist of dementia patients. And for the asthma study, 3-axes accelerometer and 3-axes gyroscope signals are acquired from the wrist and ambulatory single-lead ECG signals are acquired from the chest. These modalities have certain similarities but large varieties among themselves in terms of the kind of reaction those are capturing from the human body. Consequently, the feature extraction and engineering algorithms are also very different and independent for these modalities. In this chapter, I present my methods for extracting representative features from the wrist motion signal and the chest-worn ambulatory ECG signal. The representative features are useful for learning machine learning (ML) models of physiology and behavior from these representations. But another important use of such features can be to explore the explanatory variations and effects on these features for various physiological and behavioral states. Such exploration can enable these features to be used as biomarkers of human physiology and behavior. This kind of biomarker exploration is valuable not only for explaining the ML model mechanisms, but also for intervention design and effect simulation.

4.1 Motion Features

The wrist-worn accelerometer sensor captures directional force of the wrist motion against the gravitational force g. And the gyroscope measures the directional rotation in the three-dimensional space. The accelerometer in the Pebble watch (for the dementia study) and in the custom wrist-worn device (for the asthma study) can capture forces within ±4g range, and the gyroscope of that device captures rotation velocity within ±360 degrees per second (dps) range.

4.1.1. Data Pre-Processing

Real-world sensor signals are often noisy and suffer from missing data packets. Such signals are prone to noises from communication, physical impact, connection handover, and sensor hardware issues. To reduce the effect of such noise on the wrist motion signals, I use a data pre-processing pipeline. This pipeline consists of multiple filters and data alignment methods.

The raw sensor signals often contain random outliers, due to hardware issues or impulse current surge on device. These noises are similar to speckle noises in an image. I use a median filter of <0.25 second window, depending on the sampling rate, to remove such noises. Moreover, since human motion is mostly within the frequency range of 0.1 to 20 Hz, I apply a bandpass FIR filter with a passband from 0.1 to 20 Hz to reduce any motion artifacts. These filters are implemented using standard filtering functions provided with the Signal Processing Toolbox of MATLAB. The filtering is applied on the whole raw signal streams before performing any windowing for feature extraction.

Even after filtering, often raw sensor signals show segments of out-of-range values that are physically and electrically impossible but occur due to ADC peripherals on chip or software issues. Using simple out-of-range clippers, such regions can be identified and marked as missing data. Moreover, due to device-to-device variation, the signals vary from deployment to deployment in their calibration. Using the known range of sensors and initial non-worn segments of calibrating signals, such signals can be aligned to baseline prior to further processing.

4.1.2. Windowing

For both the BESI and the VICTER studies, wrist-worn motion sensors are used for different purposes. The variations in the objective and the settings lead to different methods for extracting the features for these applications.

For the BESI study on dementia care, the raw motion is captured by the smartwatch sampled at 50 Hz. I want to capture the dynamic variations in the wrist motion across daily activities as well as across

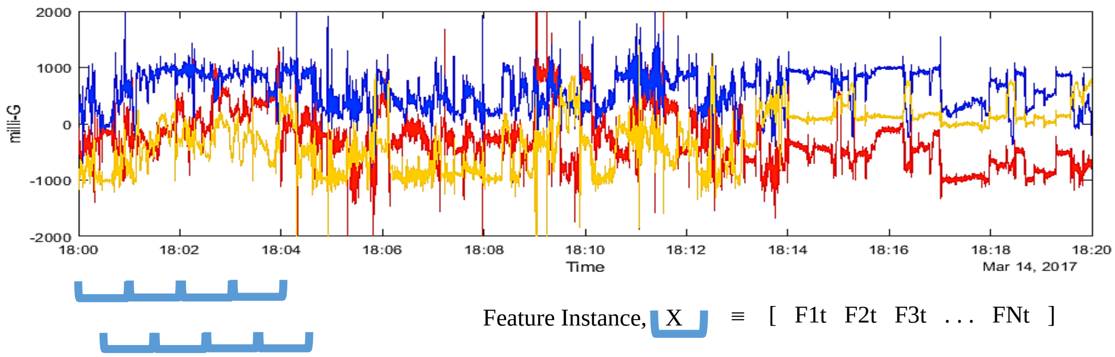


Figure 4.1: One-minute window sliding with 50% overlap for feature extraction from the wrist-worn 3-axes accelerometer signal for a 20-minute long agitation episode.

behavioral traits including agitated behaviors. Hence, a minute-window is empirically chosen for windowing. In this effort, the raw motion signal from the 3-axes accelerometers are windowed with a 60-second wide window generating 3000 samples per axis per window. This number of samples are chosen to avoid aliasing in extracting the frequency features for each axis. I slide this window with 50% overlap to acquire a three-dimensional signal window every 30 seconds. I extract feature instances from the signal for each window, i.e. every 30 seconds, for each axis (Figure 4.1).

For the VICTER project on asthma care, the raw wrist motion signal consists of 3-axes accelerometer and 3-axes gyroscope. The sensor device samples these signals at 250 Hz. This six-dimensional raw signal is windowed for feature extraction with a 15-second wide window generating 3750 samples per window per sensor axis. This number of samples are required to extract frequency features from each window without suffering from aliasing and other issues. I slide the windows with 80% overlap, which resulted in a feature vector every 3 seconds. The empirical reasoning behind this parameter is to enable instantaneous tracking of the respiratory function as fast as every 3 seconds. These streams of the 6-d signals are used to extract a feature vector instance.

4.1.3. Feature Extraction

For both applications, to represent fast and slow human motion actions with high variance predictors, the feature space is designed to contain components from statistical, frequency, and energy domains. Spatial relations among the 3 axes of the accelerometer signal are captured by calculating pairwise interaction. Frequency features include maximum and mean power spectrum density values in three frequency ranges: 0-1 Hz, 1-3 Hz, and >3 Hz, as well as zero crossing rate and mean crossing rate. Teager energy is also calculated to represent the change in energy of the signal. The interaction features are extracted across windows of pair of axes for a sensor modality. These features are presented in Table 4.1.

TABLE 4.1 PROPOSED WRIST MOTION FEATURES

Category	**Features**	**Sensor**	**Representation**	**Direction**
Statistical	Mean, Median, Maximum, RMS value, Standard deviation, Variance, Inter-quartile range	Accelerometer	The parameters of the sample distribution for the acceleration or force	Along x, y, z axes of the sensor on the wrist
		Gyroscope	The parameters of the sample distribution for the rotation velocity	Around x, y, z axes of the sensor on the wrist
Power	Mean, Maximum, and Variance of Teager energy	Accelerometer	The distribution of the amount of change in acceleration or force	Along x, y, z axes of the sensor on the wrist
		Gyroscope	The distribution of the amount of change in rotation velocity	Around x, y, z axes of the sensor on the wrist
Frequency	Mean and Maximum in the power spectrum bands for 0-1 Hz, 1-3 Hz, and >3 Hz, Zero crossing rate, Mean crossing rate	Accelerometer	Characterizing slow to fast acceleration or force of movement, frequency or repetition of force	Along x, y, z axes of the sensor on the wrist
		Gyroscope	Characterizing slow to fast rotational movement, frequency of rotation	Around x, y, z axes of the sensor on the wrist
Interaction	Mutual information, Joint entropy, Cross-correlation, Fréchet distance, Dynamic time warping	Accelerometer	Relative variation or similarity between pair of axes capturing directionality of acceleration or force	Between pairs of axes (x-y, y-z, and z-x) of the sensor
		Gyroscope	Relative variation or similarity between pair of axes capturing directionality of rotation	Between pairs of axes (x-y, y-z, and z-x) of the sensor

To the best of my knowledge, this is a novel set of features extracted from the wrist motion sensors for machine learning pipeline. Both the BESI study and the VICTER study use the same set of features for the corresponding wrist-motion sensor modalities. I acquire 18 features for each axis and 5 features for each pair of axes of a single sensor signal, accumulating to total 69 features per sensor over each window. For BESI, a feature instance of 69 values are generated every 30 seconds. For VICTER, every 3 second a feature instance is generated with 138 values. These features are used in the machine learning (ML) models to be used in application-driven model development.

4.2 Wearable ECG Features

The chest-worn wearable device acquires ECG signal from three leads at 250 Hz sampling rate. To avoid the collinearity among the signals from those three leads, I only use Lead-I signal to acquire the features. The electrical activity of the heart makes the signal vary within about one millivolt range. To reiterate, the wearable ECG signals are only collected and used for the VICTER study on asthma care.

4.2.1. Data Pre-Processing

The wearable ECG often suffers from disturbances due to baseline wondering, motion artefacts, and noise from skin contacts. Such noises are challenging and more prevalent in ambulatory and wearable ECG compared to stationary ECG. To reduce the effects of such disturbances, I first use median filtering to reduce speckle noises from skin contact or hardware issues. Then, I perform linear approximation of the baseline for each activity and detrend the signal using that approximation. Finally, to reduce effects of motion artifacts, I use a bandpass filter with 5-25 Hz pass band on the detrended signal. This preprocessing stage improves the signal quality of all the windowed wearable ECG streams.

4.2.2. Windowing

To capture stable patterns of the heart activity along with the dynamic variations across physical activities, I segment the single-lead raw wearable ECG signal using a window duration of 15 seconds. This window size is heuristically selected and may be varied across studies and sensing devices, if needed. But, to ensure temporal alignment of the wearable ECG feature instances with those from the wrist motion

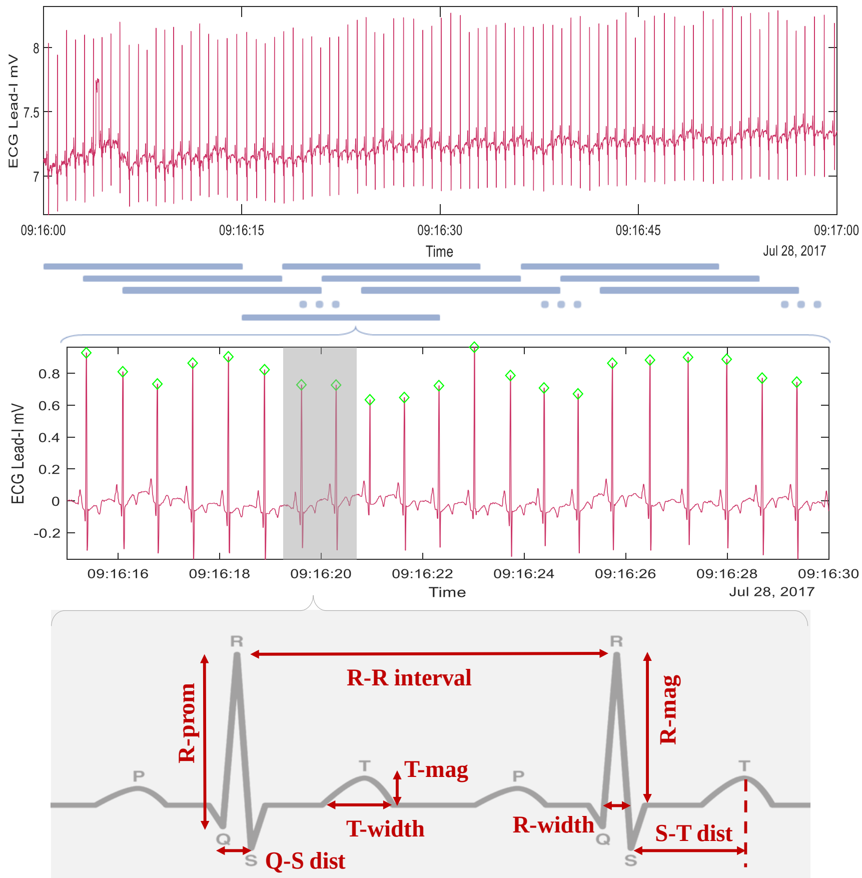


Figure 4.2: For each 15-second window, the ECG features are extracted for inferring respiratory parameters.

modality, I ensure the feature extraction to be clocked every 3 seconds, which requires us to slide the window with 80% overlap (Fig 4.2).

4.2.3. Feature Space

The wearable ECG feature space is designed to capture not only the overall characteristics of the heart's electrical activity within the time window, but also the dynamics among the individual beats within that window. Consequently, the feature space builds upon the morphological and frequency features extracted for a single beat. This endeavor first requires to identify the morphological markers for each ECG beat within the window. My approach toward achieving those markers are described here:

a. For each preprocessed signal window, I implement the standard Pan-Tompkins [141] peak detection algorithm to find the R-peak fiducial points, r = [r1, r2, …]. I use a 200 ms lockout time to avoid erroneous detection of peaks from signal noise or artifacts. Since the signal within that window is detrended or baseline adjusted in the preprocessing step, I use a local threshold for the peak detector calculated as the 70% of the highest signal value.

b. The temporal interval between consecutive detected peaks, i.e. the R-R intervals, are analyzed to identify possible missed peaks outside the ±20% deviation from the local average interval, and to update the set of peaks, r.

c. For each entry r_i in r, I search for Q, S, and T peak locations within the r_i-100 ms to r_i+500 ms segment using standard peak detection algorithm. These markers help to quantify the characteristics of each beat within the window.

Using these markers of the ECG beats, I acquire the morphological characteristics, namely the magnitude, prominence, and width of the R-wave, the magnitude and width of the T-wave, the QS distance, and the ST distance of each heart beat (Fig. 4.2.1). I also calculate the beats-per-minute (BPM) from R-R interval, and the powers, defined as the area under the triangle, of the R-wave and the T-wave. Finally, I calculate the statistical mean and standard deviation of each of these features for the individual beats within the 15 s time window to acquire the corresponding 23-dimensional feature vector instance. To the best of my knowledge, this is a novel set of morphological and power features extracted from the wearable ECG for machine learning pipeline. Every 3 seconds, the wearable ECG feature extractor sends a feature instance to the regression inference models that are used to predict respiratory parameters.

4.2.4. Data Labels

The ground truth data for the respiratory parameter, VE, are acquired asynchronously using the spirometer. I calculate the averages over 15 s window, same window length as motion and wearable ECG, to use as the response values for corresponding features. I also slide this window by 3 s, same as for the features, at each step to temporally sync the responses with the feature instances.

RELATED PUBLICATIONS:

- R Alam, et al., "Motion biomarkers for early detection of dementia-related agitation", ACM Proceedings of Workshop on Digital Biomarkers (WDB), ACM, 2017.

Chapter 5

Modeling Behavior & Physiology

Learning the behavioral and the physiological parameters from the wearable sensor signals can be achieved by building machine learning models that can infer or predict the health parameters from the raw sensor signals or the features extracted from those signals. For the two applications in hand, namely the BESI and VICTER studies, the learning goal is different. For the BESI study on dementia care, the behavior that needs to be learnt is the agitated behavior of dementia patients using the wrist motion sensor signals. For the VICTER study on asthma care, the objective is to infer the respiratory parameters, namely minute ventilation, from the wrist motion and the wearable ECG signals or the extracted features. These objectives are independent both from the response or target variables related to the application perspective and the predictor variables related to the sensor perspective. But the second and third research questions from Chapter 1 relate to both these scenarios, and I pursue to answer those questions in these contexts. Toward that goal, I explore both applications independently while trying to address the research challenges in building robust models from the collected real-world data.

5.1 Predicting Agitation in Dementia

Physical agitation is often depicted by a broad spectrum of actions and behaviors, from restlessness, repetition, pacing, and random movement to aggressive destruction, hitting, kicking, pushing, throwing, scratching, falling, hurting, and so on [87-89]. While these symptoms are often captured by the motion sensors on the patients' wrists, there is lack of knowledge about any structure of manifestation of these symptoms during an agitation episode. Moreover, such manifestation varies from person to person,

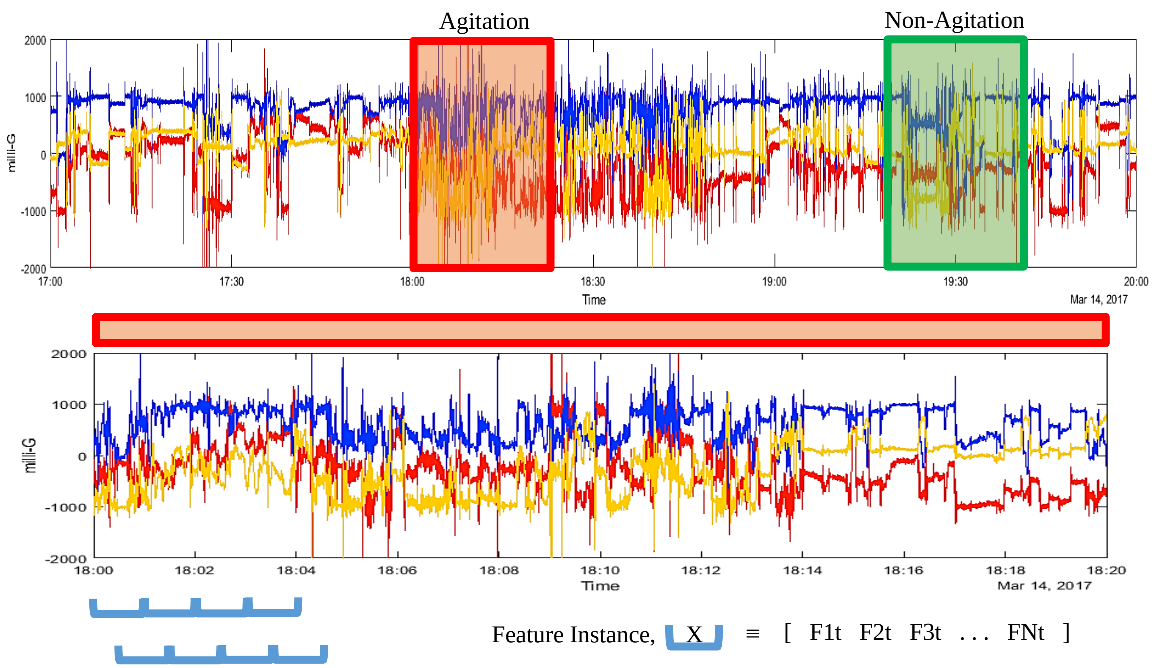


Figure 5.1: Windowing and feature extraction for an agitation episode

and even for the same person from situation to situation. Hence, to explore the usability of motion signal features for modeling physical agitation events, I attempt to find reliable motion biomarkers for detecting early signs of agitation. The feature extraction on the wrist-worn accelerometer signals transform the raw signals into a timeseries of feature instances, where each instance represents the motion features for the corresponding time window. Using these instances, I train various ML models to explore the best approach in capturing the agitated behavior of dementia patients.

Agitation is not an instant event, but more of an episode. The data labels that are acquired from the real-world users, in this case the caregivers, contain a point timestamp for that episode along with information about possible symptoms. Assuming that timestamp as the middle point of the corresponding agitation episode, I used a 20-minute window to label the agitation episode. Based on the episode label, the corresponding feature instances during that episode are labeled as agitation (Fig 5.1).

5.1.1 Learning Episodes from Instances

To capture agitation from wrist-worn accelerometer features, I explore multiple standard ML models for classification and train those to learn agitation and non-agitation classes from wrist-motion features. Three major ML models are implemented for this purpose, namely support vector machines (SVM), random forest (RF), and adaptive boosted decision trees (AdaBoost). The implementation details of these models are presented here:

5.1.1.1 Support Vector Machines (SVM)

Even for high dimensional data, the SVM attempts to classify the data by finding a hyperplane that best separates data of one class from those of other classes. And the margin of the hyperplane is identified by the support vectors that are data points closes to the hyperplane. In my implementation, the classification task is defined as a binary classification between agitation and non-agitation instances. I used a gaussian kernel with sequential minimal optimization.

5.1.1.2 Random Forest (RF)

RF classification aggregates the predictions from the ensemble of deep decision trees, each trained using N out of N instances randomly sampled with substitution from the training set. Each trained tree uses random subset of the predictors for splitting each node to avoid correlated trees in the ensemble [142]. The individual predictions of these weak learners are aggregated by majority vote or the scores averaged to get the prediction from the ensemble [143]. My implementation uses two hundred deep decision trees, while preventing overfitting to possible outliers by enforcing at-least ten observations at the leaf nodes.

5.1.1.3 Adaptive Boosted Trees (AdaBoost)

AdaBoost utilizes large number of weak learners and aggregate their output into a weighted sum to achieve the final classification [144]. This method trains learners sequentially to adaptively tweak the later learners to learn harder examples. It computes the prediction for test data as,

$$f(x) = \sum_t \alpha_t h_t(x)$$

where α_t is the learnt weight for the weak tree in the ensemble and depends on the individual classification error, and h_t is the individual prediction of the weak learner t. In my implementation, I use five hundred shallow decision trees, each with a maximum split of 5 branches with the Gini index as the split criteria.

To achieve the classification for the whole 20-minute episode from the instance outputs of these classifiers, I use majority voting on the 39 instances that belong to an episode.

5.1.2 Sequence Learning

Agitated behaviors in dementia are often considered to escalate over time, making those sequential in nature. My goal is to capture the progression of such behavior to use for preventing the escalation to any harmful health condition. To model this characteristic, I tried to formulate agitation episodes as temporally progressive action sequences. I employed recurrent neural network (RNN), which is a renowned platform for building models from sequential sensor streams. Such a network can capture the short-term temporal properties of sensor data. To model the pattern of action progression over longer duration, I incorporated long short-term memory (LSTM) cell based RNN. LSTM extends RNN with memory cells and can learn activity patterns over long durations from various streams.

Traditional neural network looks at an input data point or feature vector to model some output variables on a certain point of time. Recurrent neural network (RNN) provides a link between time steps to pass information about network states across time steps [145]. This feature makes RNN a good model for sequence modeling and implies that information from the past can be used to model a present or future

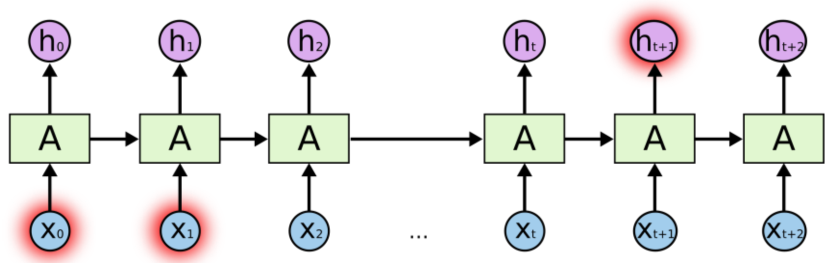
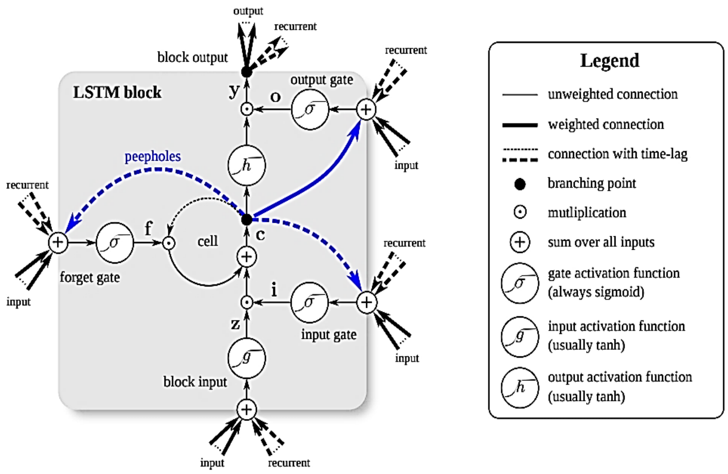


Figure 5.2: Recurrent neural network architecture

state, as shown in Fig 5.2. In this figure, each block *A* contains a neural network that models the input-output relationship at a certain time step. But, with just a simple link between two networks at two-time steps, RNNs often fail to model long-term dependencies because of issues like vanishing gradients in the back-propagation. To overcome these issues, LSTM was proposed as an extension to traditional RNN architecture to capture the dependencies across long time difference.

LSTM has been gaining momentum in sequence modeling tasks. In natural language processing applications such as machine translation, language modeling, or speech recognition, LSTM has demonstrated high potential [145]. Also, activity detection applications using wearable sensor data and using video data have employed LSTM successfully. Similar works have been done in emotion recognition from audio signals. Such applications often use single sensing modality to model the relevant activity or emotion. Motivated from the success of these works, this research aims to utilize the LSTM platform to model the relationship between context and human behavior.

LSTM introduces a “memory” cell inside each network block of RNN at each time step. This cell controls the network state propagation both from previous time step to present and from present to future time step. To achieve this control, the memory cell uses three gates: input, output, and forget gates. The basic architecture of a memory cell is shown in Fig 5.3. On each gate, there is a controller that modulates the signal past the gate. The controller signals are output of a sigmoid function (σ), which generates a signal between 0 and 1 based on the inputs to the function. The activation functions (g and h) represent two independent neural networks and generate signals between -1 and +1. The function g takes in the current input and the previous output signal to model the block input. This signal is modulated by the input gate, the controller of which depends on the current input signal, the previous output, and the *previous* cell state. The modulated signal from the input gate is used along with the modulated signal from the forget gate to update the cell memory state. The controller of the forget gate is generated the same way as the input gate controller using a different weight parameter, and this controller modulates the previous cell state to filter out the forgettable components. The current state of the cell memory is fed into the activation function h. Its output is modulated by the output gate to generate the block output using a controller that depends on



$$
\begin{aligned}
\bar{\mathbf{z}}^t &= \mathbf{W}_z\mathbf{x}^t + \mathbf{R}_z\mathbf{y}^{t-1} + \mathbf{b}_z \\
\mathbf{z}^t &= g(\bar{\mathbf{z}}^t) \\
\bar{\mathbf{i}}^t &= \mathbf{W}_i\mathbf{x}^t + \mathbf{R}_i\mathbf{y}^{t-1} + \mathbf{p}_i \odot \mathbf{c}^{t-1} + \mathbf{b}_i \\
\mathbf{i}^t &= \sigma(\bar{\mathbf{i}}^t) \\
\bar{\mathbf{f}}^t &= \mathbf{W}_f\mathbf{x}^t + \mathbf{R}_f\mathbf{y}^{t-1} + \mathbf{p}_f \odot \mathbf{c}^{t-1} + \mathbf{b}_f \\
\mathbf{f}^t &= \sigma(\bar{\mathbf{f}}^t) \\
\mathbf{c}^t &= \mathbf{z}^t \odot \mathbf{i}^t + \mathbf{c}^{t-1} \odot \mathbf{f}^t \\
\bar{\mathbf{o}}^t &= \mathbf{W}_o\mathbf{x}^t + \mathbf{R}_o\mathbf{y}^{t-1} + \mathbf{p}_o \odot \mathbf{c}^t + \mathbf{b}_o \\
\mathbf{o}^t &= \sigma(\bar{\mathbf{o}}^t) \\
\mathbf{y}^t &= h(\mathbf{c}^t) \odot \mathbf{o}^t
\end{aligned}
$$

Figure 5.3: LSTM memory cell and forward pass equations [145]

the current input signal, the previous output, and the *current* cell state. Finally, this block output is used as the recurrent signal for the *next* time step.

In this work, the LSTM neural network contains four neural layers with a shallow LSTM layer (Fig 5.4). The first layer is a sequential input layer which enables the distribution of the sequential feature vectors to be parsed. The following layer consists of a shallow LSTM layer with 64 hidden units. The internal state of each LSTM cell is forwarded to the next cell, and only the last cell outputs the classification of the sequence as either agitation or non-agitation. A fully connected layer is used to facilitate the acquisition of the outcome based on learned gating schemes. Finally, a SoftMax layer is incorporated for output unit activation in either agitation or non-agitation categories. Each sequence comprises of 39 feature instances, where each feature instance is 33 dimensional. For training, 10 sequences constitute a mini-batch, and the training is

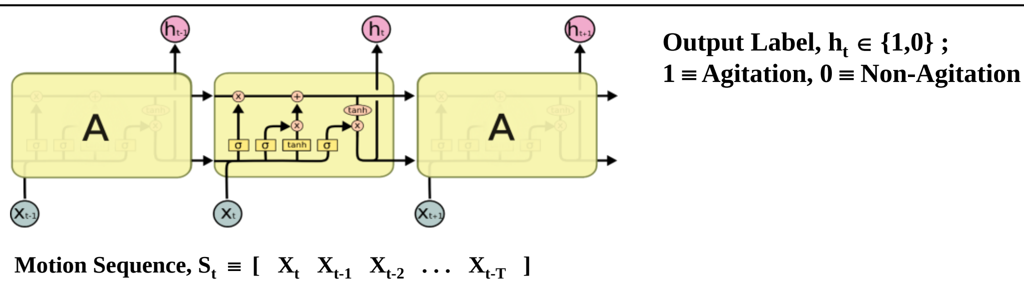


Figure 5.4: LSTM implementation

run for maximum 20 epochs. 75% of the agitation and non-agitation episodes are randomly chosen for training this model with 5-fold cross-validation. The rest 25% random holdout dataset was used for performance evaluation.

5.1.3 Learning with Sparse Labels

Agitation is highly unpredictable and sparse in occurrence without any established pattern or cause, which challenges the established pattern learning approaches. Moreover, acquiring the annotations or labels for such behavior in the real-world is challenging, often due to privacy and compliance related concerns. With the dementia patient case study over the month-long BESI deployments, no camera was used, and the labels were actively provided by the caregivers when they observed any agitated behavior of the patients. To reduce burden and improve compliance, the caregivers could label an episode by simply pressing a button on their smart watches, which stored that timestamp as a marker for the agitation episode, or used a tablet device to report the event on a custom Android app. In this process, the lack of any label is assumed to be non-agitation by default. Yet, the reliability or precision of such user-provided labels may not be guaranteed, there can be time lags between the actual time of occurrence and the label timestamps.

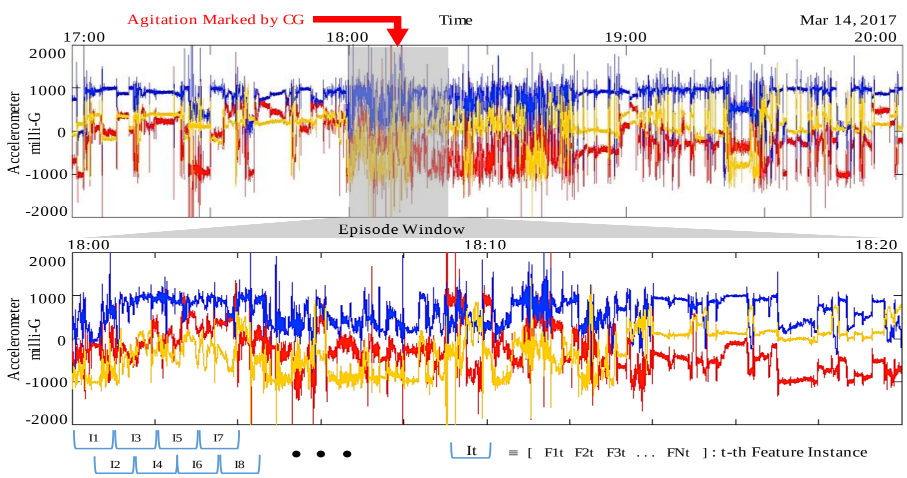


Figure 5.5: Wrist motion signal from the patient smart watch during an agitation episode, labeled by the caregiver with a single timestamp. Feature instances are extracted from a window surrounding that timestamp, but all instances in that window may not be representative of the behavior.

To address these issues, I first try to improve the model performance from the possible imprecision in the labels. Now, there is no established method to judge the reliability of such labels. And, removing unreliable labeled data may worsen the situation by reducing the size of the training set. Hence, I tried to reduce the impact of such imprecision on the model performance by adding flexibility in characterizing the feature space. Multiple instance learning (MIL) provides an appropriate framework for this task [146-148]. Multiple instance learning (MIL) deals with training data arranged in sets, called bags. Supervision is provided only for entire sets, and the individual label of the instances contained in the bags are not provided (Fig 5.6). To utilize such flexibility, I implement two MIL models, namely bag-level kernel-based support vector machine (MI-SVM) [149] and boosting bag-level decision stumps (MIL-Boost) [150]. The implementation details and the evaluation process are described here.

5.1.3.1 Bag of Instances

For the physical agitation models, every 30 seconds, a 69-dimensional feature instance vector is acquired from the motion signals. In contrast, the patient behavior is not continuously labelled, only the sparse agitation episodes are marked. To address this sparsity, I start with the marked timestamps (T) and assume the agitation episodes to have some components within [T-10, T+10], i.e. 20 minutes surrounding those markers. Each of these episode windows provides 39 of the feature instances. Instead of assigning the same "positive" labels for all these instances, I compose a "bag" of these instances and assign the label to the bag. Then, each "positive" bag should have at-least one instance representing the agitated behavior. Similarly, I randomly pick timestamps distant from the markers and compose "negative" bags, guaranteed to have no "positive" instances. With this formulation, these bags are used toward training MIL models for

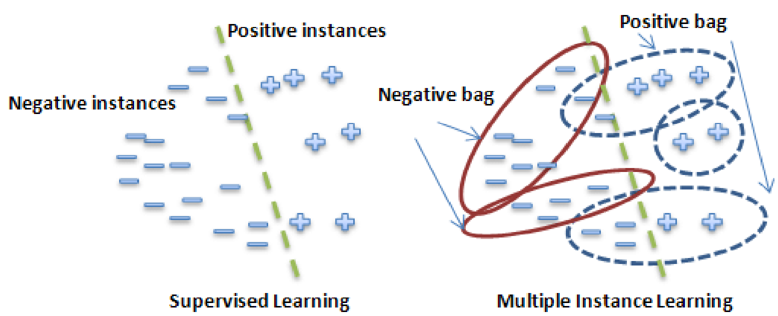


Figure 5.6: Standard assumption of MIL increases the flexibility on label precision

predicting physical agitation episodes.

5.1.3.2 MIL Models

I use these bags of instances to train two variants of MIL models, namely bag-level kernel-based support vector machine (MI-SVM) [9] and boosting bag-level decision stumps (MIL-Boost) [10], as described here. Each feature instance is represented by the row vector $\mathbf{x}_i = [x^{(1)}, x^{(2)}, \dots x^{(d)}]_i$, d is the number of features, and a bag is presented as a set of instances $X = \{\mathbf{x}_i\}$, for i = 1, 2, … n; n is the size of the bag. A classifier defined on the instance space $\mathfrak{X}$ can be written as $f: \mathfrak{X} \rightarrow \{0,1\}$, and a bag classifier as $g(X) \in \{0,1\}$.

The standard MIL assumes that every positive bag contains at-least one positive instance and no negative bag contain any positive instance. To estimate an instance-level classifier f (xi; θ), MIL models maximize the number of positive bags that contain at least one instance in the hyperplane θ and the number of negative bags that do not project on θ. Then the bag classifier can be implemented as the max rule:

$$g(X) = \max_{\mathbf{x}_i \in X} f(\mathbf{x}_i); \text{where } f(\mathbf{x}_i; \theta) = \begin{cases} 1 & \text{if } \mathbf{x}_i \in \theta \\ 0 & \text{otherwise} \end{cases}$$

Such models learn the instance level classifier from the bag labels by an iterative process of finding nearby instances.

The MI-SVM model improves the f (xi; θ) using kernel-based similarity measurements from the hyperplane. It attempts to maximize a functional margin of a bag from the hyperplane, defined as:

$$\gamma_X \equiv Y_X \max_{\mathbf{x}_i \in X}(\langle w, \mathbf{x}_i \rangle + b)$$

Here, Y_X is the label for the bag X. This formulation ensures that the margin separates the "least negative" instances and captures up to the "most positive" instance.

MIL-Boost trains many weak decision stumps from the instances, ft(xi) instead of a single model. The classifications of these stumps are used to find the probability of the bag classification by the "noisy-OR" formulation, as shown here:

$$f(\mathbf{x}_i) = \sum_t \alpha_t f_t(\mathbf{x}_i) \qquad \text{[Boosting]}$$

$$p_{\mathbf{x}_i} = \Pr\big(f(\mathbf{x}_i)\big) = 1/\big(1 + e^{-f(\mathbf{x}_i)}\big) \qquad \text{[Logistic]}$$

$$p_X = \Pr(g(X)) = 1 - \prod_i (1 - p_{\mathbf{x}_i}) \qquad \text{[Noisy OR]}$$

Each round of boosting searches for a weak classifier that maximizes the likelihood, and estimates the parameter αt. These models are trained to predict physical agitation on the bag of instances, which constituted the sequence in modeling the LSTM model before. I used the MIL toolbox for Matlab [16] to implement these models.

5.1.4 Contextual Ensemble (ConxEns) Classification

With an aim to improve the model performance while ensuring generalizability, I propose a novel contextual ensemble pipeline for hierarchical classification using additional source of information beyond data labels. This pipeline is called ConxEns. For the BESI study, the caregivers not only mark the timestamps of the agitation events, they also provide observations about the symptoms of that event. Such symptoms can be used to group the agitation events. The distribution of agitation event frequencies varies among these groups, same as the variation of symptoms across patients.

The pipeline incorporates a group of banks of the classification models as described in Section 5.1.1; each model learnt to predict agitation with a specific symptom from wrist motion features (Fig 5.7). Each bank can be dedicated for one or more symptom, and is independent of other banks. A bank facilitates the modularity to use various symptom-specific models, depending on the application. The models within a bank are also independent of each other and operate in parallel. For my implementation to predict agitation, I trained fourteen groups of models, each group corresponding to a symptom presented in Fig 2.1. For each symptom in the group, I implement the AdaBoost and train it independent of other models for other banks.

Then, this pipeline uses a symptom classifier to learn the symptom from the wrist motion signals. The symptom classifier is built to determine the associated observation of symptoms by using the wrist motion features. This classifier can be trained independently of the symptomatic classification models, hence, can be updated over time with possible changes in the symptom space or the predictor feature space. This modular design also enables the classifier to be transferable across applications. In this implementation, I train an AdaBoost as the symptom classifier. My implementation of AdaBoost uses shallow decision trees with maximum five splits. I design the decision trees to use Gini's diversity index as the metric for node splitting. I enforce choosing the split predictor based on chi-squared tests of independence not only between each predictor and the response, but also between each pair of predictors

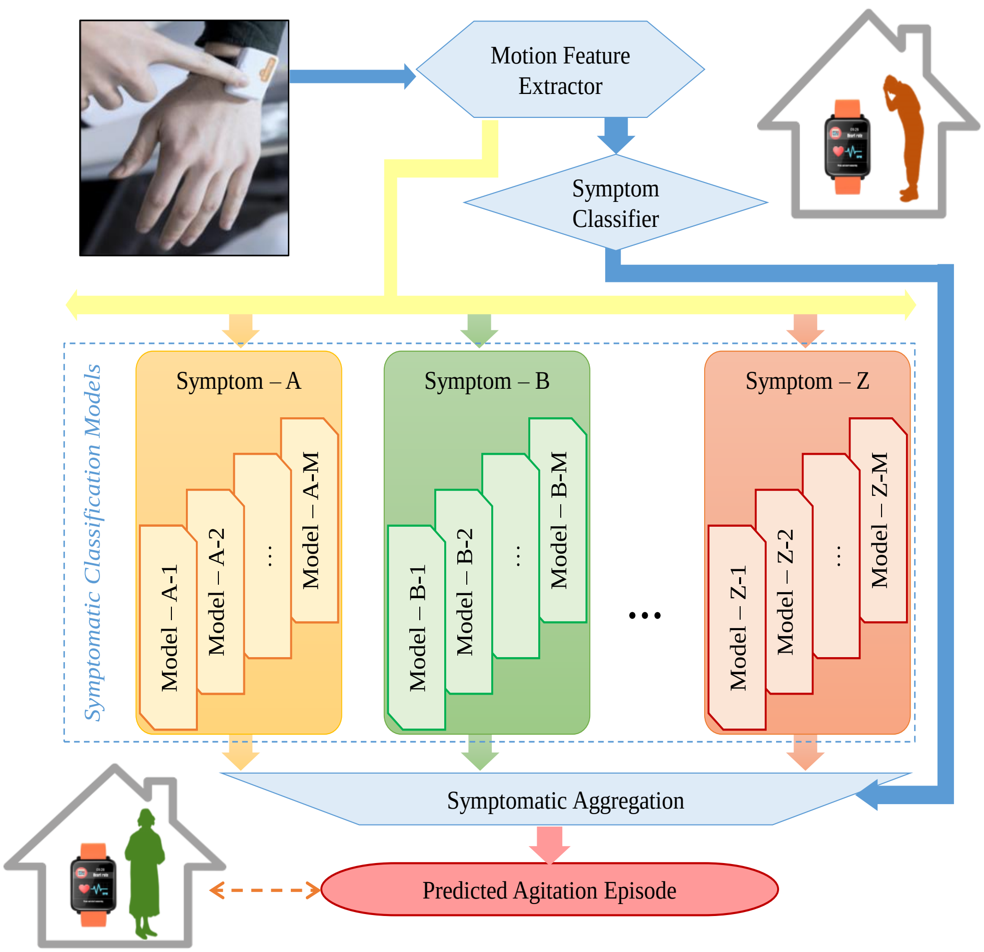


Figure 5.7: Proposed ConxEns pipeline for classification: Symptoms classified from wrist motion is used to aggregate agitation markers independently predicted from same sensor signal.

and the response. For the AdaBoost, I trained the ensemble with an upper bound of two hundred iterations. For a motion feature instance z_i, the model yields the posterior probabilities $\mathbf{p}_i = [p^{(1)}, p^{(2)}, \ldots p^{(m)}, \ldots p^{(M)}]_i$, where M is the total number of symptoms, and $p_i^{(m)} \in [0,1]$ is the posterior probability of that instance being in symptom m such that $\sum_m p_i^{(m)} = 1$.

The final step of the ConxEns pipeline is an ensembler or aggregator that combines the predictions from the symptomatic classification banks based on the output of the symptom classifier. This proposed pipeline incorporates a novel conditional aggregation method. Based on the level of the posterior probability for the classified symptom of an instance, multiple options can be applied, from selecting a symptomatic classifier bank to using weighted averaging of the classification scores. The instance-wise classification of ConxEns are aggregated by majority voting to achieve a classification for the episode.

This implementation of ConxEns pipeline for predicting agitated behavior in dementia from wrist motion signals. To demonstrate the generalizability of this pipeline for other health applications, I also implement ConxEns for inferring respiratory parameters from the wrist motion and wearable ECG signals as part of the VICTER study, as will be presented later.

5.2 Inferring Respiration for Asthma Care

With the motivation toward asthma attack prevention, I attempt to estimate the respiratory parameters, namely minute ventilation (V_E), using wearable ECG and wrist-worn accelerometer and gyroscope sensors. Challenges toward this objective span from sensor noise reduction to physiological signal representation and modeling the relationship between sensor data and respiration. The inter-personal and contextual variations among the physiological parameters challenge the exploration of interpretable relationships.

5.2.1. Respiration Regression Models

First, I attempt to model VE from the wearable ECG signal extracted features alone. Then, I also try to model VE from wrist motion signal extracted features. These two independent modeling approaches use some established ML modeling framework, namely generalized linear models (GLM), random forest

(RF), support vector machines (SVM), Gaussian process regression (GPR), and neighborhood component analysis (NCA). The implementations of the models are described below. These implementations demonstrate the methodology to learn physiological parameters using ML methods.

For the explanation of the model functionalities, let, each sensor feature instance be represented by the row vector $\mathbf{x}_i = [x^{(1)}, x^{(2)}, \ldots x^{(d)}]_i$, as d is the number of features. Also, let the corresponding respiratory parameter be represented by the scalar y_i, for i = 1, 2, … n; n is the size of the training set. In the minute ventilation implementations, y_i represents V_E. Each test instance feature vector is represented by $\mathbf{x}_t$ and the predicted respiration value as $\hat{y}$.

5.2.1.1. Generalized Linear Model (GLM)

GLM extends linear regression by allowing for exponential distributions of the prediction error. While linear regression models the response variables to linearly vary with predictors, in GLM, a link function of the distribution mean of the response is expected to vary linearly with predictors [151-153]. Assuming an exponential distribution for the response, the link function f of the distribution mean μ is modeled against the feature instances $\mathbf{x}_i$'s using coefficient set $\beta = [\beta_0, \beta_1, \ldots \beta_d]$. This is formulated as $E[y] = \mu = f^{-1}(\beta\mathbf{X})$. The parameter β can be constrained using the elastic net with regularization parameter λ and scaling factor α. Thus, elastic net drives some coefficients to zero and reduce dimensionality [49,50], by minimizing the cost function, L(β), defined using the deviance of the model fit:

$$L(\beta) = \frac{1}{n}\text{Deviance}(\beta) + \lambda\frac{(1-\alpha)}{2}\|\beta\|_2^2 + \lambda\alpha\|\beta\|_1$$

My implementation of GLM uses the identity function as the link function f, assuming normal distribution for the respiration parameter. The coefficient β is learnt from the sensor features to model the distribution mean of the respiratory parameter. I dynamically adjust the regularization by calculating λ from the training sample size. I combine both L^1 and L^2 penalties on β using α = 0.5.

5.2.1.2. Random Forest (RF)

RF regression aggregates the predictions from the ensemble of deep decision trees, each trained using N out of N instances randomly sampled with substitution from the training set. Each trained tree uses random subset of the predictors for splitting each node to avoid correlated trees in the ensemble [142-143].

The individual inferences of these weak learners are averaged to get the prediction from the ensemble. Permuting one predictor at a time, out-of-bag losses are analyzed to rank the predictors based on their contribution on the prediction.

For each implementation, I employ an ensemble of two hundred decision trees to learn the respiratory parameter from the ECG feature set. These trained trees are designed to grow deep, while preventing overfitting to possible outliers by enforcing at-least ten observations at the leaf nodes. The mean squared error is used as the split criterion for these regression trees. To avoid the bias in the predictor selection at each node split, I address the interactions between the predictors by using interaction test. This method conducts chi-square tests of independence between each predictor and the response, as well as between each pair of predictors and the response. I prioritize the predictor that minimizes the p-values for both tests.

5.2.1.3. Support Vector Machine (SVM)

SVM regression uses a kernel-based transformation of the feature space, and learns an optimal hyperplane that limits the prediction error within an "insensitivity" threshold ε. The hyperplane is characterized by the support vectors, and is learnt as the coefficients $\alpha = [\alpha_1, \alpha_2, \ldots \alpha_n]$ and bias b by minimizing the loss function, L(α), defined in [154] as:

$$L(\alpha) = \frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)G(\mathbf{x}_i, \mathbf{x}_j) + \varepsilon\sum_{i=1}^{n}(\alpha_i + \alpha_i^*) - \sum_{i=1}^{n} y_i(\alpha_i - \alpha_i^*)$$

under constraints on α_i's using the box constraint C. The optimization process is also constrained by the Karush-Kuhn-Tucker complementarity conditions. Using a Gaussian kernel for the transformation, a new feature instance is used to predict the corresponding response as:

$$\hat{y} = f(\mathbf{x}_t) = \sum_i(\alpha_i - \alpha_i^*)\, G(\mathbf{x}_i, \mathbf{x}_t) + b;\; G(\mathbf{x}_j, \mathbf{x}_k) = e^{-\|\mathbf{x}_j - \mathbf{x}_k\|^2}$$

My implementations of the SVM regression dynamically adjusts the value of ε based on the distribution of the response variables. The box constraints are similarly adjusted as ten times that of ε. I use the sequential minimal optimization as the algorithm for minimizing the cost function with 10^{-6} feasibility gap as the associated convergence criterion.

5.2.1.4. Gaussian Process Regression (GPR)

GPR is a non-parametric kernel-based approach. In this probabilistic method, the response, y_i, is explained using a latent function of the predictors, $f(x_i)$, along with the linear combination of a transformation $h(x_i)$ of the predictor space [155]:

$$P(y_i|f(\mathbf{x}_i), \mathbf{x}_i) \sim \mathcal{N}(h(\mathbf{x}_i)^T\beta + f(\mathbf{x}_i), \sigma^2)$$

Here, the basis function h is a transformation of the feature space, chosen empirically. The linear combination coefficient vector β, the latent function f, and the noise variance, σ^2 is learnt from the data. The latent variables, $f_i = f(\mathbf{x}_i)$, are assumed to possess a Gaussian process prior, such that for all variables, $\mathbf{f} = [f_1, f_2, \dots f_n]$, we get $P(\mathbf{f}|\mathbf{x}_1, \mathbf{x}_2, \dots \mathbf{x}_n) \sim \mathcal{N}(0, K)$. For this prior, K is the covariance matrix defined using the kernel function k, as $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$. The parameter, θ, associated with the choice of the kernel function is learnt during training. Using estimated β, θ, and σ^2, the latent variable $\hat{f} = f(\mathbf{x}_t)$ is inferred for any test instance $\mathbf{x}_t$. The joint GP prior $P(\hat{f}, \mathbf{f})$ is used with the likelihood for $\mathbf{y} = [y_1, y_2, \dots y_n]$, which is $P(\mathbf{y}|\mathbf{f})$, to get the joint posterior:

$$P(\hat{f}, \mathbf{f}|\mathbf{y}) = \frac{P(\hat{f}, \mathbf{f})P(\mathbf{y}|\mathbf{f})}{P(\mathbf{y})}$$

where $P(\hat{f}, \mathbf{f}) \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K_{\mathbf{f},\mathbf{f}} & K_{\hat{f},\mathbf{f}} \\ K_{\mathbf{f},\hat{f}} & K_{\hat{f},\hat{f}} \end{bmatrix}\right)$ and $P(\mathbf{y}|\mathbf{f}) \sim \mathcal{N}(\mathbf{h}^T\beta + \mathbf{f}, \sigma^2 I)$. I marginalize this posterior over **f** to acquire $\hat{f}$, which is used to get the response $\hat{y}$, i.e. the respiratory parameter.

Depending on the kernel function k, the covariance matrix K captures the similarity among feature instances. Parameters of these kernel functions are the signal variance, σ_s^2, and the characteristic length scale, σ_l^2. Automatic relevance determination (ARD) uses different length scale parameter σ_r^2 for each feature r = 1, 2, … d, to investigate their individual contribution in inferring the latent and the response variables [156].

I implement GPR and ARD by choosing the Matern kernel function k with separate σ_r^2 for each feature, defined as:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \sigma_s^2\left(1 + \sqrt{3m}\right) \exp\left(-\sqrt{3m}\right);\ m = \sum_{r=1}^{d} \frac{(x_i^{(r)} - x_j^{(r)})^2}{\sigma_r^2}$$

Here, the parameters, $\theta = [\sigma_s^2, \sigma_r^2]$, are learnt during training. To avoid local minima, I initialize σ_s^2 using the variance of the response variable, and σ_r^2 using feature variances. The noise variance σ^2 is initialized similarly as σ_s^2. For transforming the feature space, I use linear basis function $h(\mathbf{x}_i) = [1\ \mathbf{x}_i]$, and learn the coefficients β from the data.

5.2.1.5. Neighborhood Component Analysis (NCA)

NCA is non-parametric and avoids any assumption about the sample distribution. It uses a stochastic neighbor selection rule to assign any test instance the response value of its selected neighbor. This rule reduces its dependence on the amount of training data and the risk of overfitting. NCA attempts to learn a quadratic distance metric, representable as linear transformation to low-dimensional input space, minimizing the regression loss [157,158]. For any $\mathbf{x}_s$ in training set S and a test instance $\mathbf{x}_t$, distance metric D_w is defined using predictor weights, w_r, as,

$$D_w(\mathbf{x}_t, \mathbf{x}_s) = \sum_{r=1}^{d} w_r^2\ |\mathbf{x}_{tr} - \mathbf{x}_{sr}|$$

Then, the stochastic selection uses the probability of any $\mathbf{x}_s$ in S being the nearest neighbor of $\mathbf{x}_t$ as p_{ts}:

$$p_{ts} = P(\text{neigh}(\mathbf{x}_t) = \mathbf{x}_s | S) = \frac{\exp\left(-\|D_w(\mathbf{x}_t, \mathbf{x}_s)\|\right)}{\sum_{s \in S} \exp\left(-\|D_w(\mathbf{x}_t, \mathbf{x}_s)\|\right)}$$

Using the response of the nearest neighbor relative to the learnt distance metric, the test response $\hat{y}$ is inferred.

My implementation uses mean absolute error as the metric for measuring the regression loss, and learns the distance metric using the limited memory Broyden-Fletcher-Goldfarb-Shanno algorithm. Instead of storing all the training samples, I store the linear transformations. For each context-bank, I dynamically choose the regularization parameter based on the size of the training set for that bank.

The above described implementations capture the decision making regarding the selection of the hyperparameters for learning physiological parameters from sensor signals.

5.2.2. Contextual Ensemble (ConxEns) Regression

After implementing these models independently for the wearable ECG and the wrist motion sensor features, I aim to improve the performance as well as the generalizability by proposing a novel contextual inference pipeline, named ConxEns. With this pipeline, I explore the value of context in modeling respiration and understanding the modeled relationships. Context can often be scavenged from external sources including wearable sensing modalities. But, incorporating such context into the respiration estimation may yield specialized models lacking flexibility and generalization.

The pipeline incorporates a group of banks of the regression models described above; each model learnt to infer contextual respiration from wearable ECG features only. Each bank can be dedicated for one or more context, and is independent of other banks. A bank facilitates the modularity to use various context-specific models, depending on the application. The models within a bank are also independent of

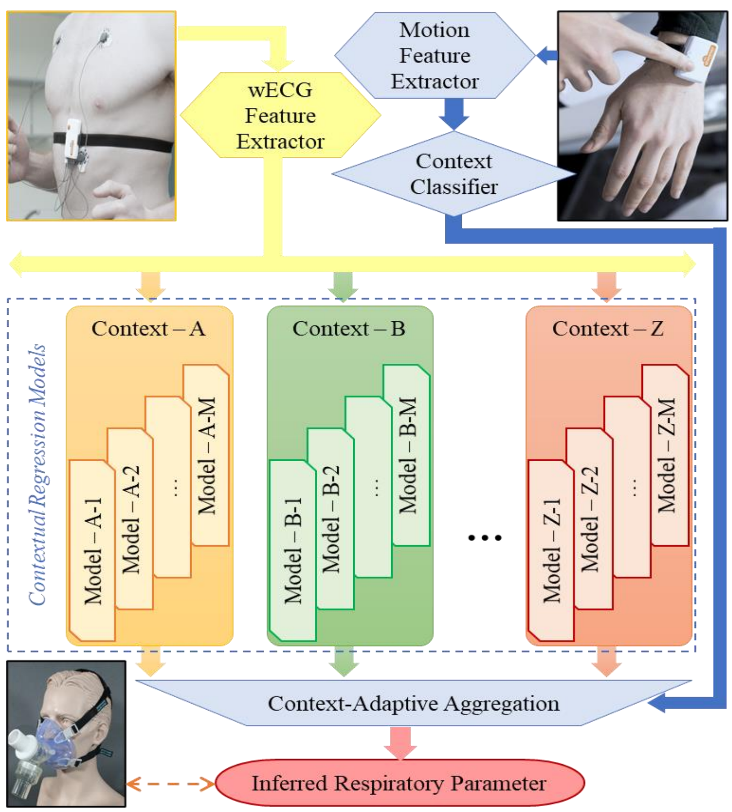


Figure 5.8: Proposed ConxEns pipeline for regression: Context classified from wrist motion is used to aggregate respiratory parameters independently inferred from wearable ECG.

each other and operate in parallel (Fig. 5.8). For my implementation to infer V_E, I trained separate groups of banks of models. Each model in a context-bank is trained and operates independent of other models in that bank and those in other banks.

Then, this pipeline uses a context classifier to learn the context from the wrist motion signals. The inference from the respiration regression models are aggregated as a probabilistic ensemble with a prior acquired from the context classifier. The implementation details of the context classifier and the aggregator are presented here:

5.2.2.1. Context Classifier

The context classifier is built to determine current physical activity context by using the wrist motion features. This classifier can be trained independently of the regression models, hence, can be updated over time with possible changes in the context space or the predictor feature space. This modular design also enables the classifier to be transferable across applications. In our implementations, we train a single instance of the classifier and use that instance in both BR and VE inference pipelines.

I design the classifier as an ensemble of shallow decision trees using the totally corrective boosting algorithm, known as TotalBoost [46]. Unlike other boosting algorithms, this method updates the weight distribution for the "hard" examples in the training set by finding the distribution with minimum relative entropy to the initial distribution. In [159], this relative entropy is expressed as, $\Delta(d,d_0) = \sum_i d_i \ln(d_i/d_{0i})$, the KL divergence of two distributions. This algorithm prioritizes hypotheses that maximize the minimal margin of classification and minimize the number of observations below that margin, thus guarantees low generalization error.

My implementation of TotalBoost uses shallow decision trees with maximum five splits. I design the decision trees to use Gini's diversity index as the metric for node splitting. I enforce choosing the split predictor based on chi-squared tests of independence not only between each predictor and the response, but also between each pair of predictors and the response. For the TotalBoost, I trained the ensemble with an upper bound of two hundred iterations. A margin precision parameter of $v = 0.01$ is used as a constraint in updating the hypothesis with respect to all past hypotheses. For a motion feature instance z_i, the model yields the posterior probabilities $\mathbf{p}_i = [p^{(1)}, p^{(2)}, \dots p^{(m)}, \dots p^{(M)}]_i$, where M is the total number of contexts, and $p_i^{(m)} \in [0,1]$ is the posterior probability of that instance being in context m such that $\sum_m p_i^{(m)} = 1$.

5.2.2.2. Context-Conditioned Aggregator

The final step of the ConxEns pipeline is a probabilistic ensembler or aggregator that combines the inferences from the contextual regression banks based on the output of the context classifier. Traditional regression aggregators attempt either to select the best performing regressor from a group or to average their performances to achieve overall better performance [160]. This proposed pipeline incorporates a novel conditional aggregation method merging both strategies depending on the context. Based on the level of the posterior probability for the classified context of an instance, multiple options can be applied, from selecting a regression bank to using weighted averaging of the regression inferences from the contextual banks and to build a joint distribution of the respiration parameter and the context. The posterior probabilities for each context can be used as the averaging weights or as a prior to the joint distribution. The joint distribution can be constructed as:

$$P(y, z|x) = P(y|z, x).P(z|x)$$

where z is an additional source of information, in this case the activity context.

I implemented the ConxEns pipeline for inferring V_E from wrist motion and wearable ECG signals. To demonstrate the generalizability of this pipeline for other health parameter regression, I also implement ConxEns for inferring respiratory rate or breathing rate (BR) from the wrist motion and wearable ECG signals as part of the VICTER study.

As presented before, the proposed ConxEns platform has been implemented for both behavior classification and physiology regression tasks from wearable sensor signals. This novel methodology is proposed to answer the research questions 2 and 3 regarding the modeling approach of human parameters from sensor signals while optimizing performance and generalizability.

RELATED PUBLICATIONS:

- R Alam, et al., “Inferring physical agitation in dementia using smartwatch and sequential behavior models”, IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), IEEE, 2018.

- R Alam, et al., “Multiple-instance learning for sparse behavior modeling from wearables: Toward dementia-related agitation prediction”, IEEE Annual International Conf of the Engineering in Medicine and Biology Society (EMBC), IEEE, 2019.
- R Alam, et al., “Non-invasive inference of minute ventilation using wearable ECG and Gaussian process regression”, IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), IEEE, 2019.
- R Alam, et al., “Inferring respiratory minute volume from wrist motion”, IEEE Annual International Conf of the Engineering in Medicine and Biology Society (EMBC'19), IEEE, 2019.
- R. Alam, et al., “Wearable respiration monitoring: Interpretable inference with context and sensor biomarkers”, IEEE Journal of Biomedical & Health Informatics, Under review, 2020.

Chapter 6

Model Performance

The proposed and implemented methods for modeling behavior and physiology with wearable sensor signals from the real-world settings are evaluated for both predictive performance and generalizability. The two health monitoring applications, that are used in this research as the case studies, pose two different kinds of problem: the agitated behavior prediction for dementia care is a classification problem, and the respiratory parameter inference for asthma care is a regression problem. Consequently, the model performance evaluation requires different sets of metrics.

6.1 Agitated Behavior Prediction Models

Using data collected from 12 dementia patients for over 12 months duration at their residential spaces, I explore multiple ML models to learn the agitated behavior of the patients from their wrist-worn accelerometer sensor. The dataset contains 571 agitation events marked by the caregiver of the patients, each event corresponding to a 20-minute episode. Out of the 12 months long dataset, only about 1200 minutes data belongs to the 'agitation' class. The rest of the dataset is assumed to be of the 'non-agitation' class. This is a heavily imbalanced dataset and can skew the model performance evaluation. To reduce the skewness, I randomly sampled from the 'non-agitation' episodes pool to rebalance the dataset to 1:4 ratio between agitation and non-agitation classes.

The rebalanced dataset is divided into training and testing sets using 30% hold-out for testing with stratification to ensure similarity of class distributions in these two subsets. The training set contains 70% of the dataset and are used to train the models in 5-folds cross validation.

For a classification task, one of the most used metrics for evaluating model performance is accuracy. Accuracy is defined as the ratio of the sum of correct predictions to the number of total samples,

$$\mathrm{Acc} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} = \frac{\mathrm{N_{true}}}{\mathrm{N_{total}}}$$

where N_{true} is the number of samples classified correctly and N_{total} is the total number of samples. Accuracy is a good metric for evaluating models trained and tested on balanced dataset, but for datasets with imbalanced class distribution, this metric fails to capture the model performance. For example, if the test dataset has high number of samples from class c, then the model will have high accuracy by classifying all samples as belonging to that class.

When the data are imbalanced by high margins, model performance needs to be evaluated using metrics that are independent of class distribution. Two such metrics are weighted f1-score (F_w) and mean f1-score (F_m), as defined as

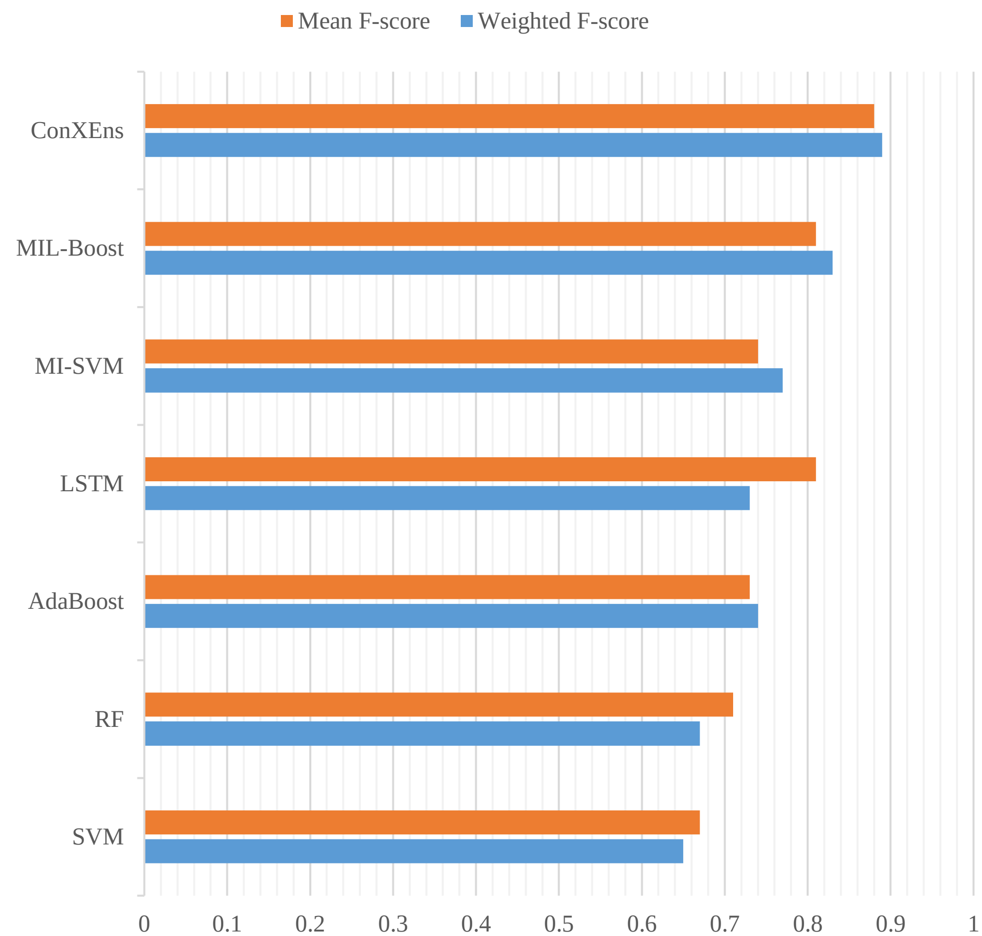
$$\mathrm{F_w} = 2\sum_{\mathrm{c}} \frac{\mathrm{N_c}}{\mathrm{N_{total}}} \frac{\mathrm{precision_c} \times \mathrm{recall_c}}{\mathrm{precision_c} + \mathrm{recall_c}},$$

and

$$\mathrm{F_m} = \frac{2}{|\mathrm{c}|}\sum_{\mathrm{c}} \frac{\mathrm{precision_c} \times \mathrm{recall_c}}{\mathrm{precision_c} + \mathrm{recall_c}},$$

where, N_c and N_{total} are number of samples in class c and in total respectively, and $|c|$ is the total number of classes. These metrics are better than only accuracy, precision, or recall, as these are normalized against class distribution.

While both the mean (F_m) and the class-weighted (F_w) f1-scores of the predictions are good indicators of model performance against class imbalance, the application in hand may want to prioritize or compare performance for certain class compared to others. For example, the application may require higher performance for identifying the ‘positive’ class than sacrificing some performance for the ‘negative’ class, i.e. to achieve higher true positive rate than lower false positive rate. The receiver-operating-characteristic (ROC) curves are useful tools to compare the performances and trade-offs of the models, especially in this



	SVM	RF	AdaBoost	LSTM	MI-SVM	MIL-Boost	ConXEns
Mean F-score	0.67	0.71	0.73	0.81	0.74	0.81	0.88
Weighted F-score	0.65	0.67	0.74	0.73	0.77	0.83	0.89

Figure 6.1: Performance comparison of the proposed models for predicting agitation in dementia.

kind of application-driven decision making. The ROC curve is defined by false-positive rate (FPR) and true-positive rate (TPR) as x and y axes, respectively. This curve can be used to observe the relative trade-offs between true-positive and false positive. This curve is also known as the sensitivity vs (1 − specificity) plot, as TPR is equivalent to sensitivity and FPR to (1 – specificity). The ROC curve is constructed by using each prediction result or instance of a confusion matrix as a point on the curve.

Using these metrics, the proposed models are evaluated on the 30% hold-out test dataset. The performance in shown in Fig 6.1 with accompanying value table. As a baseline, three popular supervised learners, namely support vector machine (SVM), random forest with bagged trees, and random under-sampling based AdaBoost classifiers are evaluated.

From the result, it can be observed that the sequential model using LSTM outperforms the instance-based baseline learning models by at-least 8% in mean F-scores. It is notable that for an imbalanced data with about 80% non-agitation and 20% agitation events, most of the traditional learners achieve good accuracy by classifying any instance as the majority class. On the other hand, the F-score metrics are more representative of the confusion matrix and reduces the effects of data imbalance. LSTM captures the temporal relationship among the minute actions of the behavioral events and can learn various patterns for such behaviors. Thus, it classifies based on sequence of instances rather than a cluster of instances, and consequently performs better than the baseline methods.

By looking closer, it can be noted that the LSTM model is performing better in only learning the positive class i.e. agitation, but not much improvement in preventing false positives. Rather, the AdaBoost has shown better performance in reducing false positives. To reduce the effect of possible noisy labels in the data, I implement and evaluate the multi-instance learning-based models MI-SVM and MIL-Boost. The performance of these models is also shown in Fig 6.1.1. The result shows the effectiveness of these models in overcoming the challenges associated with imprecise labels from real-world users. The MIL-Boost improves the weighted F-score by 10%.

Moreover, in Fig 6.2, I evaluate the receiver-operating-characteristic (ROC) curves to compare the

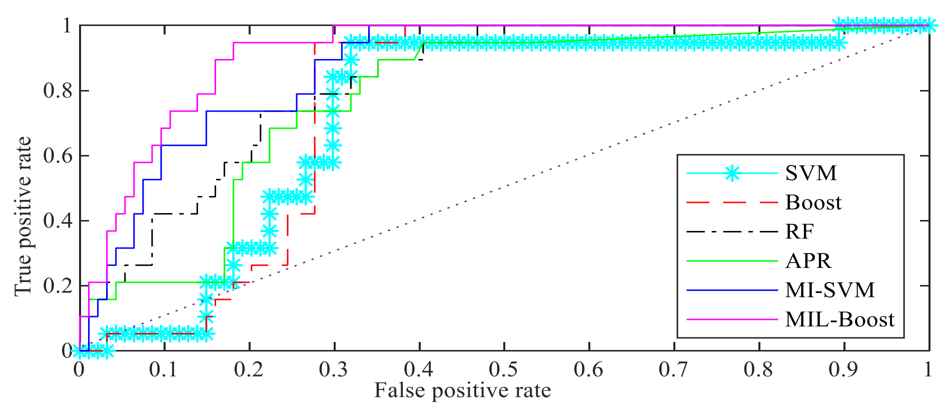


Figure 6.2: ROC curve shows the ratio of TPR to FPR across the learnt models of agitation in dementia.

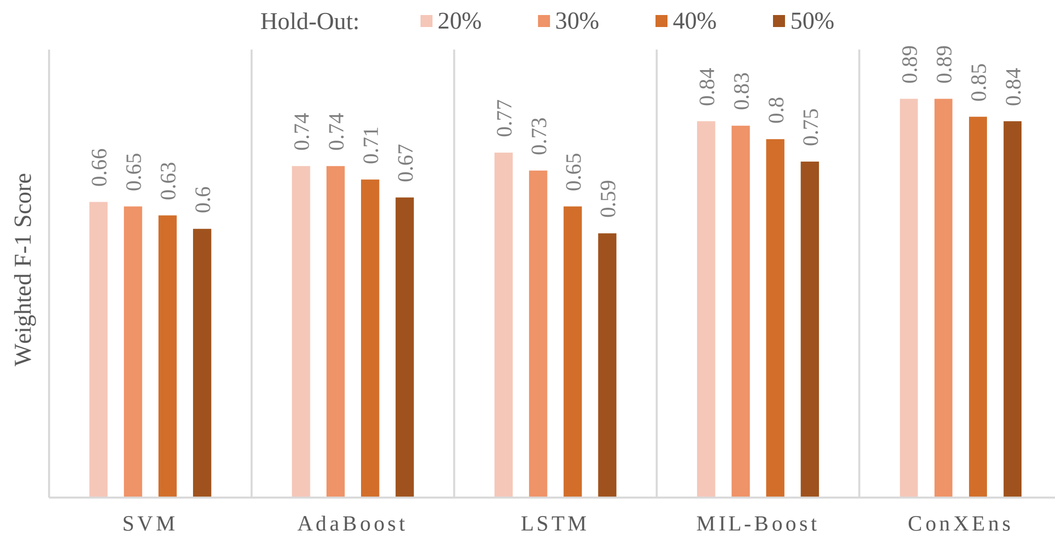


Figure 6.3: Generalizability across various hold-out test ratios for the models of agitation in dementia.

performances and trade-offs of the models, especially in reducing false-positives. The ROC curves and the weighted F-scores demonstrate the improvement in the models' specificity. MI-SVM shows better performance compared to single-instance SVM, and MIL-Boost improves both the positive and the negative predictive values. The ROC curves show the improved positive likelihood ratio of the MIL classifiers.

To explore the generalizability of the proposed models, especially to times beyond the training set, I evaluate the models the same way for multiple train-test hold-out ratios: 80%-20%, 70%-30%, 60%-40%, and 50%-50%. The standard approach of leave-one-subject-out is not a suitable method in this context, as the number of agitation episodes as well as the severity of those episodes vary a lot from patient to patient which leads to the absence of any similarity among the distribution of agitation events across patients. The result of these evaluations is shown in Fig 6.3 using the weighted f1-score of the model predictions on the test sets. As shown in this figure, with smaller data available for training, the performances drop for all the models. But some models appear to be more robust than others across this generalization evaluation. The proposed ConxEns modeling pipeline only loses 5% performance; it shows F_w score of 89% when trained on 80% data and tested on the rest 20%, whereas it achieves 84% score when trained on 50% data and tested on the rest. On the other hand, LSTM loses predictive performance in F_w score from 77% to 59% as training set is reduced from 80% to 50%. The instance-based implementation of AdaBoost also demonstrates robustness across different training-testing ratios.

This result demonstrates the improved performance and the generalizability of the proposed ConxEns pipeline in learning the agitated behavior of dementia patients from their wrist motion sensor signals.

6.2 Respiration Inference Models

Data collected from 15 healthy subjects, each performing a physical exercise protocol yielding about 15 minutes of sensor and respiration data for the five physical activities: rest, walk, run, bike, and wave (excluding rest periods between activities), are used in training and evaluating the regression models for inferring respiration from wrist motion and wearable ECG sensors. The preprocessing generates 16 instances per minute, totaling about 3450 samples of data.

The dataset is divided into 70%-30% training-testing hold-out ratios with stratification for all activity context. The training set is used to train the regression models for inferring minute ventilation (V_E) from wearable ECG signals and wrist motion signals. The training process uses a 5-folds cross validation.

For a regression task, the evaluation metrics for comparing the performance of the trained models are mostly related to the residual i.e. the difference between the regressed value and the true value. Various forms of residual-based metrics are popular in practice, such mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE). MSE is defined as the sample mean of the square of the residuals,

$$\mathrm{MSE} = \frac{1}{n}\sum_{n}(y - \hat{y})^2$$

where n is the sample size and $\hat{y}$ is the inferred value. RMSE is the square root of the MSE metric. Using similar notations, MAE is defined as,

$$\mathrm{MAE} = \frac{1}{n}\sum_{n}|y - \hat{y}|$$

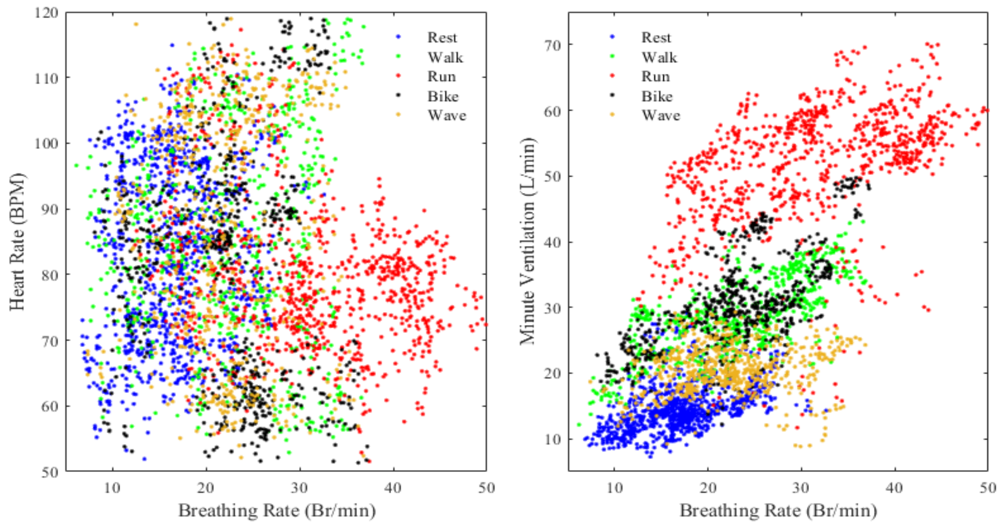


Figure 6.4: Variations in breathing rate against heart rate and against minute ventilation from 15 subjects.

As seen from the definitions, MSE has a squared penalty for outliers. In the respiration inference task, the distribution of respiratory parameters varies widely across activities and subjects (Fig 6.4). The inter-personal and contextual variations among the physiological parameters challenge the exploration of interpretable relationships. With a goal to capture those variations, I adopt the MAE as the evaluation metric for this task.

I built independent models for inferring V_E from the wearable ECG and the wrist motion sensor data separately to explore the relationship between V_E and the sensor signals. A set of models, namely generalized linear regression (GLM), random forest (RF), support vector machines (SVM), and Gaussian process regression (GPR), is learnt to model V_E from wearable ECG. And another set of the same models is

TABLE 6.1 Performance comparison across sensors and model selection

Models	**Chest-worn Wearable ECG**		**Wrist-Worn Accelerometer & Gyroscope**	
	RMSE	**MAE**	**RMSE**	**MAE**
GLM	12.55	9.65	16.56	13.63
RF	11.57	9.01	11.76	7.97
SVM	10.68	7.87	6.62	4.76
GPR	8.23	6.01	5.19	3.62

learnt from wrist-worn accelerometer and gyroscope. Using the respective feature instances presented in Chapter 4, these models infer V_E for each feature instance. These models are trained on the 70% training dataset previously mentioned with 5-folds cross validation. Their performances in inferring V_E are shown in Table 6.1 using the root mean squared error (RMSE) and the mean absolute error (MAE) of regression. The unit for this residual is L/min. As shown in Table 6.1, when inferring independently, wrist-worn sensor feature-based models perform better than chest-worn ECG feature-based models. These performances correspond to the effect of ambulatory noises on the sensor data and the trained models.

To improve the performance of such models, I propose a contextual inference pipeline, ConxEns. This pipeline combines both chest-worn and wrist-worn modalities in a hierarchical manner. I evaluate the proposed pipeline for context classification, and contextual respiration inference for V_E inference. To demonstrate the generalizability of this approach to other health parameters, I also infer breathing rate (BR) from these sensors.

Moreover, to demonstrate the generalizability of the implemented methods, we conduct performance evaluation over a range of train-test hold-out percentages from 80% training - 20% testing to 70-30, 60-40, 50-50, 40-60, 30-70, and 20-80 percentages. The result of each stage of the pipeline is presented below with demonstration of this generalizability evaluation.

6.2.1 Context Classification

For both BR and VE inference pipelines, we use the same context classifier that identifies the physical activities from the wrist-worn motion sensor-based features. This classifier is implemented as a total-boost classifier. The performance of this module is evaluated as classification task. For each train-test ratio, we use the hold-out test set to evaluate the trained classifier with metrics such as accuracy, true positive rate (TPR), and false negative rate (FNR). The resulting scores and the confusion matrices for four ratios are shown in Fig. 6.5. Since the dataset is not imbalanced, I use accuracy as the metric for evaluating the classification task. The TPR and the FNR are presented to demonstrate the trade-off between the activity classes that the models are conducting. Over all the train-test ratios, the mean accuracy is 99.66% with a range from 99.5% to 99.9%. For any context across the train-test ratios, the lowest TPR is 98.4% and the highest FPR is 1.6%. Even when trained on only 20% of the data, the model accuracy is 99.5%, with a lowest TPR 98.8% and a highest FNR is 1.2%. This result shows both the robustness and the generalizability of the classifier, even when trained on only 20% and tested on the rest of the data.

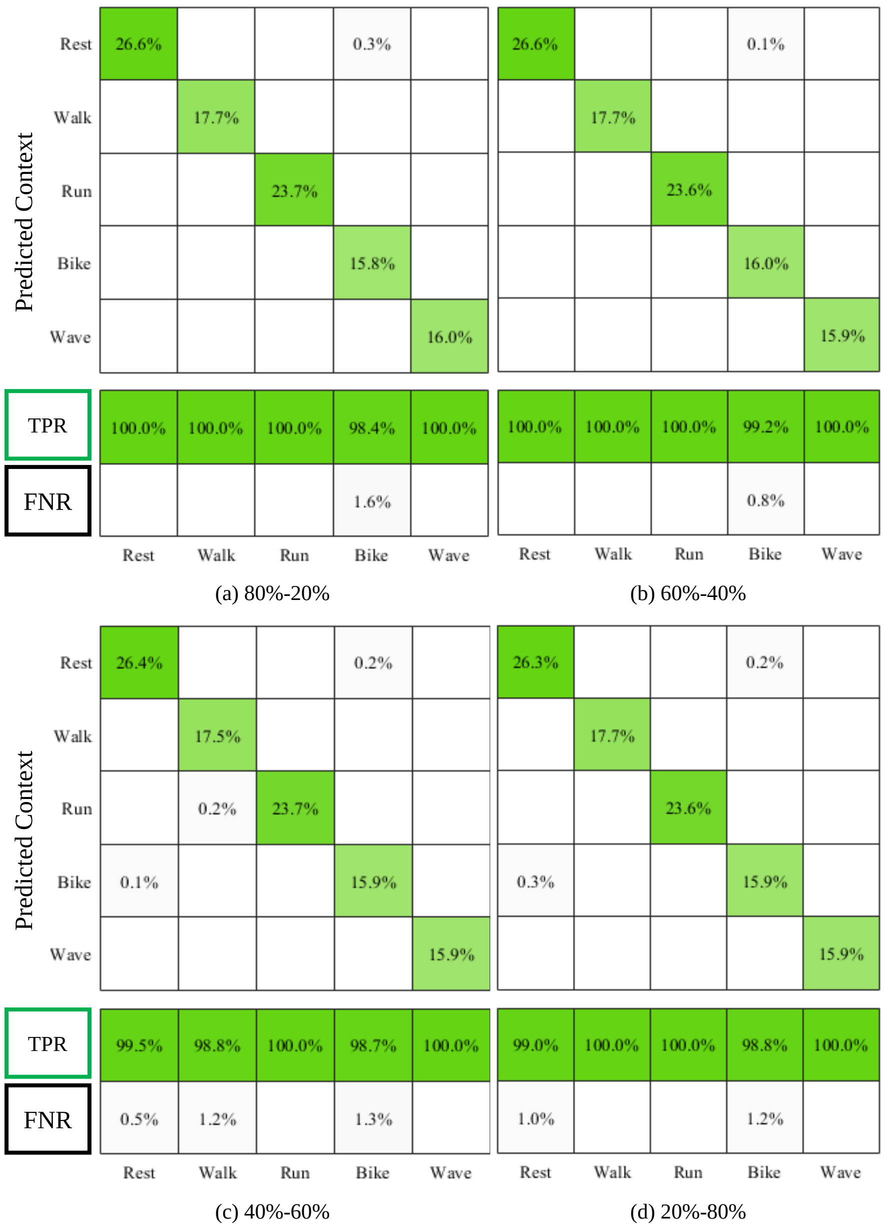


Figure 6.5: Context classification performance over varying train-test hold-out percentages: training with (a) 80% data, (b) 60%, (c) 40%, and (d) 20% only, yet maintaining ≥99.5% accuracy.

6.2.2 Respiration Inference

Using the ECG features and the contextual pipeline, I built independent pipelines to infer two respiratory parameters, BR and V_E. For this regression task, the inference loss is evaluated using mean absolute error (MAE) as shown in Fig 6.6. For BR, this loss is calculated in breaths per minute (Br/min), and V_E in liters per minute (L/min). This figure shows the results for the evaluation with 70% training and 30% hold-out test data. For this evaluation, the best performance, for both BR and VE, is acquired with NCA as the contextual regression model in the implemented pipelines. Here, for BR inference, the mean loss over all activities is 1.17 Br/min; including 0.7 Br/min during rest to 1.39 Br/min during run. And, for V_E, the overall loss is 1.39 L/min, with 0.87 L/min at rest and 1.87 L/min during run. Similarly, overall losses for using GPR contextual models are, respectively, 1.32 Br/min and 1.46 L/min; and, for using SVM, 1.59 Br/min and 1.75 L/min. These losses are notably lower, for similar contexts, compared to existing solutions, even with stationary ECG, as well as those presented in Table 6.1. This result demonstrates the value of the novel wearable ECG features in capturing the physiological relationship.

The performance comparisons between the context-agnostic and the proposed contextual models

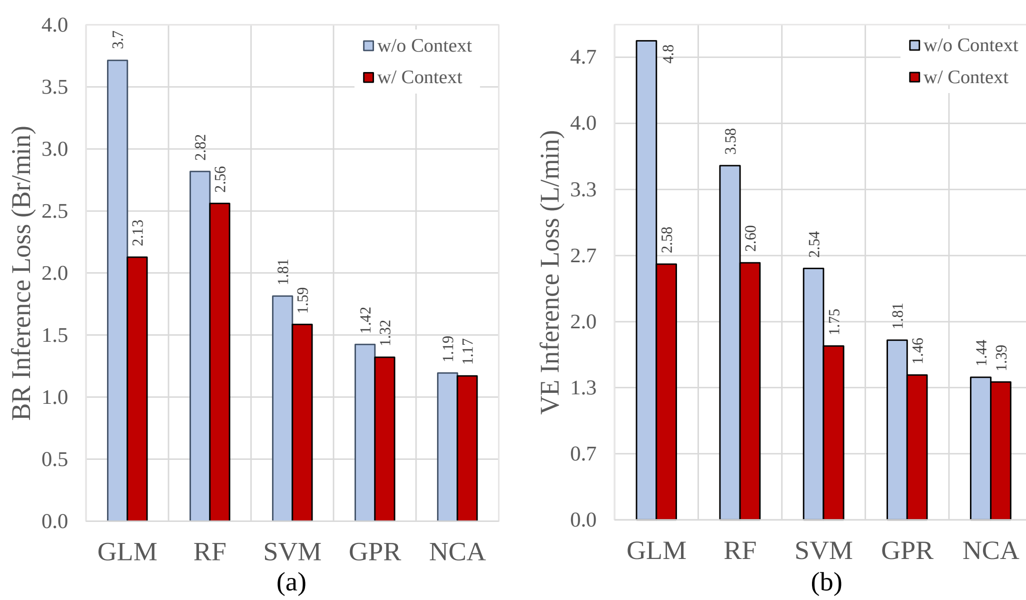


Figure 6.6: MAE inference loss of the proposed contextual pipeline with different regression models and context-agnostic models of same kind: (a) breathing rate and (b) minute ventilation.

are also notable in Fig. 6.6. A context-agnostic model is an implementation of the same kind of regression model trained without the context data. For the 70%-30% evaluation, the contextual pipeline outperforms context-agnostic models for every choice of regression model.

To evaluate robustness and generalizability, we conduct this analysis across multiple train-test percentages; the result is presented in Fig. 6.7. In this figure, downward arrows refer to loss reductions, i.e. performance improvements. Unsurprisingly, the inference performance slightly worsens with the reduction in training data. But, differences between context-agnostic and contextual models remain steady across the spectrum. For light-weight models (not required to store all samples or transformations) like GLM, RF, and SVM, the performance improves more dominantly than for neighborhood-based heavy-weight (need to store the training set) models such as GPR and NCA. Moreover, the impact of context is higher for inferring VE than for BR, as the arrows are longer for Fig. 6.7 (b), highlighting the effect of volumetric variations.

These performances across varying train-test ratios answer the research questions 2 and 3 about modeling health parameters with high performance and generalizability from real-world sensor data.

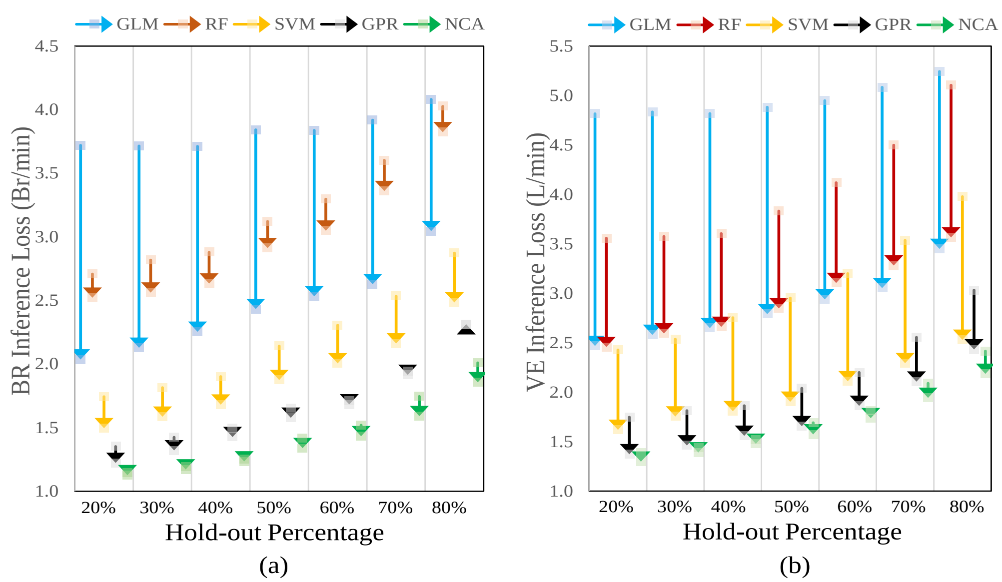


Fig. 6.7. Generalization and performance improvement (magnitude and direction, shown as arrows) in inference of (a) BR and (b) VE across varying train-test hold-out percentages.

Chapter 7

Conclusion

This research is motivated by the possibilities of the real-world impact that sensing and ML together can bring to every aspects of human welfare, especially healthcare. Toward that future, the sensor-ML systems need to generalize well to real-world unknown and unpredictable scenarios. Recent scrutiny of ML and AI systems in their applicability to healthcare has shown both promises and challenges, including areas such as privacy, security, bias, ethics, as well as toward personalization, generalization, explainability, and trust [12, 161, 162]. Among those, a critical challenge for the current state of research is to learn models from real-world noisy sparse-labeled data. In this research, I try to address the challenge of learning health parameters in an efficient and generalized manner from raw sensor signals within the scope of wearable health monitoring applications.

For two major healthcare applications, one on dementia care and another on asthma care, I use wearable sensors to continuously monitor the patients' wrist motion and chest ECG. I build models of their behavior and physiology using those sensor signals to predict application-driven health parameters. For the dementia study, I build models of agitated behavior in dementia patients using labels of such events provided by their caregiver. And for the asthma study, I model the respiratory parameters using gold standard data collected from Spirometers. To ensure robust operation in the real-world, I aim to make the models robust against real-world scenarios.

Toward that aim, I propose the following hypothesis and establish it using data-driven evidences acquired from the above-mentioned health monitoring applications:

A holistic approach from fault tolerance in sensor systems to

contextual association of data-driven models improves

the robustness of sensing-driven models of

human behavior and physiology in real-world settings.

While the common knowledge about modeling tells us that more data helps model performance, this hypothesis looks at the orthogonal direction for the meaning of 'more'. Instead of using more samples or data points or more sensor modalities, I try to achieve robust performance by incorporating supplemental information beyond the application-specific labels about the available data samples. While many applications may not have availability of such information, for real-world health monitoring, contextual information has always been regarded valuable, and maybe easily achievable from the user or the study design. Especially when the health application-specific parameters are often labeled by real-world users, some additional information regarding those labels can be achieved with less effort. For example, when a user provides a label for an eating event during any meal, she can easily answer a question about where or what she is eating. The hypothesis claims such information can improve the performance for models aimed toward detecting eating events. With this hypothesis, for the dementia care study, I try to improve the performance and generalizability of the model of agitated behavior in dementia patients by incorporating caregiver provided observation about the symptoms of agitation events. Similarly, for the asthma study, I learn the physical activity context of the users from the study protocol, and I try to use that information in improving model robustness. To explore this hypothesis for real-world applications, it requires both a robust sensing system to acquire data for long periods continuously from real people and a method to utilize the real-world data without reliable continuous labels for model improvement. In this work, I present my approaches toward achieving these goals.

I propose a novel wearable-edge ML platform, BESI, that facilitates in-home smart watch-based human sensing without restrictions on companion phone, any active interactions, or user circadian routine. Using robust monitoring and recovery strategies, BESI reduces the burden of long-term operation maintenance. I implemented a network of edge devices in home settings and leveraged their computational resources as well as their geophysical location to overcome some of the limitations of smart watch-based sensing. BESI has been deployed in real patients' residences for months and successfully acquired more than 95% data from these deployments. These data are used toward evaluating the research claim.

I build a data-driven modeling pipeline for predicting physical agitation from dementia patients' wrist motion signals. Using the BESI data, I extract representative features of the motion signal and train supervised sequential models. The model learns to detect episodes of physical agitation with about 0.8 F-1 score for the positive class, but often triggers false alarms during non-agitation periods. To address the challenges with noisy data labels from real-world users, I evaluate MIL-based models to achieve slight improvement of the performance. Similarly, for the asthma study, I design models of the respiratory parameters form users' wrist motion and wearable ECG signals using standard ML models.

For these two modeling goals for real-world health monitoring applications, I try to evaluate the above hypothesis. I propose a novel contextual ensemble pipeline, called ConxEns, for both classification and regression tasks. I implement the ConxEns for both the agitation and the respiration models. The cores of these implementations are utilization of standard ML models with a proposed probabilistic ensembler to achieve ensembles of contextual model outcomes. These implementations of ConxEns demonstrate both performance improvement and generalization. Compared to the baseline models of the agitated behaviors, this approach improves performance by about 20% even when trained on much less data. Similarly, the performance of the ConxEns implementation for respiration inference outperforms baseline models by reducing inference loss by about 50% even when trained on little amount of data. These outcomes support the hypothesis and highlights the need for multidimensional information related to the data labels.

With this research, I have proposed a novel sensor-ML network to acquire wearable sensor data in real-world residential settings and evaluated this platform in real deployments. I have developed and published baseline models for learning the agitated behavior of dementia patients from their wrist acceleration signals. I have also developed and published baseline models for inferring respiratory parameters from wrist-worn motion and chest-worn ECG sensors. I have proposed a novel contextual ensemble learning pipeline and evaluated this platform for both classification and regression tasks associated with the dementia care and asthma care applications.

One of the limitations of this work is that the proposed system and the proposed method has only been evaluated in small number of applications. Patient dynamics vary from application to application, and the human-factor effect of these variations are not explored in this research. Another limitation is that the proposed model is not evaluated for an end-to-end learning task and for aggregating deep learning-based models. Moreover, the effect of label noise and real-time performance are also not explored.

The outcomes of this research encourage future endeavors in multiple directions. Beyond the scope of wearable sensing modality, the value of supplemental information for in-situ sensing or for electronic health records can be explored. Such research can lead to robust model performance for applications ranging from smart home health monitoring and smart city healthcare research to personalized healthcare decision making or prescription suggestion. Similar research directions can be explored in the healthcare applications domain, for out-of-hospital health monitoring for multitudes of chronic disease patients and beyond, as well as to personalized health context design in a city or community-based healthcare model. From the model design perspective, research on interpretable and explainable ML/AI models can be benefitted from similar approach of hierarchical multi-task learning ensembles. Such pipelines can identify the model failure scenarios as well as explain outcomes from contextual perspectives. In healthcare applications, this kind of explainable approach can be useful for designing intervention as well as building personalized health solutions.

References

[1] Center for Disease Control and Prevention, “Health and economic costs of chronic diseases,” online, https://www.cdc.gov/chronicdisease/about/costs/, accessed on: Aug 27,2020.

[2] Zwar, Nicholas, Mark Harris, Rhonda Griffiths, Martin Roland, Sarah Dennis, Gawaine Powell Davies, and Iqbal Hasan. "A systematic review of chronic disease management." (2017).

[3] G. Smith et al., “The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death,” Resuscitation, vol. 84, no. 4, 2013.

[4] Bernell, Stephanie, and Steven W. Howard. “Use your words carefully: What is a chronic disease?” *Frontiers in public health* 4 (2016): 159.

[5] U. Varshney, “Pervasive healthcare and wireless health monitoring,” Mobile Networks and Applications, vol. 12, no. 2-3, 2007, pp. 113-127.

[6] H. Alemdar, et al, “Wireless sensor networks for healthcare: A survey,” Computer Networks, vol. 54(15), 2010.

[7] S. Islam, D. Kwak, M. Kabir, M. Hossain, and K. Kwak, “The internet of things for health care: A comprehensive survey,” IEEE Access 3, 2015, pp. 678-708.

[8] K. Peetoom, et al., “Literature review on monitoring technologies and their outcomes in independently living elderly people,” Disability and Rehabilitation: Assistive Technology, vol. 10, no. 4, 2015, 271-294.

[9] A. Pantelopoulos, and N. Bourbakis, “A survey on wearable sensor-based systems for health monitoring and prognosis,” IEEE Transactions on Systems, Man, and Cybernetics, Part C, vol. 40, no. 1, 2010, pp. 1-12.

[10] N. Krishnan, et al., “Activity recognition on streaming sensor data,” Pervasive and Mobile Computing 10, 2014.

[11] P. Rashidi, and A. Mihailidis, “A survey on ambient-assisted living tools for older adults,” IEEE Journal of Biomedical and Health Informatics, vol. 17, no. 3, 2013, pp. 579-590.

[12] R. Deo, “Machine learning in medicine,” Circulation 132, no. 20, 2015, pp. 1920-1930.

[13] S. Russell, and P. Norvig, “Artificial intelligence: A modern approach,” Pearson Education Ltd, 2016.

[14] T. Davenport, et al., “Artificial intelligence for the real world,” Harvard Business Review 96, no. 1, 2018.

[15] M. Jordan, “Artificial intelligence - The revolution hasn’t happened yet,” Harvard Data Science Review, 2019.

[16] M. Jordan, and T. Mitchell, “Machine learning: Trends, perspectives, and prospects,” Science 349, 2015.

[17] M. Chan, D. EstèVe, J. Fourniols, C. Escriba, and E. Campo. "Smart wearable systems: Current status and future challenges." Artificial Intelligence in Medicine 56, no. 3, 2012, pp. 137-156.

[18] J. Beaudin, S. Intille, and E. Tapia, “Lessons learned using ubiquitous sensors for data collection in real homes,” CHI'04 Human Factors in Computing Systems, ACM, 2004, pp. 1359-1362.

[19] J. Bardram, T. Hansen, M. Mogensen, and M. Soegaard, “Experiences from real-world deployment of context-aware technologies in a hospital environment,” Intern Conf on Ubiquitous Computing, 2006, pp. 369-386.

[20] C. D. Kidd, R. Orr, G. D. Abowd, C. G. Atkeson, I. A. Essa, B. MacIntyre, et al., “The aware home: a living laboratory for ubiquitous computing research,” International Workshop on Cooperative Buildings, Springer Berlin Heidelberg, 1999, pp. 191-198.

[21] E. M. Tapia, S. S. Intille, and K. Larson, “Activity recognition in the home using simple and ubiquitous sensors,” International Conference on Pervasive Computing, Springer Berlin Heidelberg, 2004, pp. 158-175.
[22] A. Bamis, D. Lymberopoulos, T. Teixeira, and A. Savvides, “The BehaviorScope framework for enabling ambient assisted living,” Personal and Ubiquitous Computing, vol. 14, no. 6, 2010, pp. 473-487.
[23] M. Rowe, S. Lane, and C. Phipps, “CareWatch: a home monitoring system for use in homes of persons with cognitive impairment,” Topics in Geriatric Rehabilitation, vol. 23, no. 1, 2007, pp. 3.
[24] D. F. Mahoney, E. L. Mahoney, and E. Liss, “AT EASE: automated technology for elder assessment, safety, and environmental monitoring,” Gerontechnology, vol. 8, no. 1, 2009, pp. 11-25.
[25] S. Junnila, H. Kailanto, J. Merilahti, A.-M. Vainio, A. Vehkaoja, M. Zakrzewski, et al., “Wireless, multipurpose in-home health monitoring platform: two case trials,” IEEE Transactions on Information Technology in Biomedicine, vol. 14, no. 2, 2010, pp. 447-455.
[26] A. D. Wood, J. A. Stankovic, G. Virone, L. Selavo, Z. He, Q. Cao, et al., “Context-aware wireless sensor networks for assisted living and residential monitoring,” IEEE Net., vol. 22, no. 4, 2008, pp. 26-33.
[27] D. I. Tapia, A. Abraham, J. M. Corchado, and R. S. Alonso, “Agents and ambient intelligence: case studies,” Journal of Ambient Intelligence and Humanized Computing, vol. 1, no. 2, 2010, pp. 85-93.
[28] G. Sacco, V. Joumier, D. Nelly, D. Arnaud, A. Derreumaux, J. H. Lee, et al., “Detection of activities of daily living impairment in Alzheimer's disease and mild cognitive impairment using information and communication technology,” Clinical interventions in aging, 2012, no. 7, pp. 539-549.
[29] D. J. Cook, “Health monitoring and assistance to support aging in place,” Jour. of Univ. Computer Science, vol. 12, no. 1, 2006, pp. 15-29.
[30] A. Lotfi, C. Langensiepen, S. M. Mahmoud, and M. J. Akhlaghinia, “Smart homes for the elderly dementia sufferers: identification and prediction of abnormal behaviour,” Journal of Ambient Intelligence and Humanized Computing, vol. 3, no. 3, 2012, pp. 205-218.
[31] C. Dixon, R. Mahajan, S. Agarwal, A. J. Brush, B. Lee, S. Saroiu, et al., “An operating system for the home,” Presented as part of the 9th USENIX Symposium on Networked Systems Design and Implementation (NSDI 12), 2012, pp. 337-352.
[32] F. Viani, F. Robol, A. Polo, P. Rocca, G. Oliveri, and A. Massa, “Wireless architectures for heterogeneous sensing in smart home applications: concepts and real implementation,” Proceedings of the IEEE, vol. 101, no. 11, 2013, pp. 2381-2396.
[33] C. W. Mundt, K. N. Montgomery, U. E. Udoh, V. N. Barker, G. C. Thonier, A. M. Tellier, et al., “A multiparameter wearable physiologic monitoring system for space and terrestrial applications,” IEEE Trans. on Info. Tech. in Biomed., vol. 9, no. 3, 2005, pp. 382-391.
[34] M. Sung, C. Marci, and A. Pentland, “Wearable feedback systems for rehabilitation,” Journal of Neuroengineering and Rehabilitation, vol. 2, no. 1, 2005, pp. 1.
[35] D. Malan, T. Fulford-Jones, M. Welsh, and S. Moulton, “CodeBlue: an ad hoc sensor network infrastructure for emergency medical care,” Int. Wor. on Wearable and Implantable Body Sensor Networks, vol. 5, 2004.
[36] U. Anliker, J. A. Ward, P. Lukowicz, G. Troster, F. Dolveck, M. Baer, et al., “AMON: a wearable multiparameter medical monitoring and alert system,” IEEE Transactions on Information Technology in Biomedicine, vol. 8, no. 4, 2004, pp. 415-427.
[37] C. R. Baker, K. Armijo, S. Belka, M. Benhabib, V. Bhargava, N. Burkhart, et al., “Wireless sensor networks for home health care,” 21st Int. Conf. on Advanced Information Networking and Applications Workshops, vol. 2, IEEE, 2007 pp. 832-837.
[38] D. M. Taub, S. B. Leeb, E. C. Lupton, R. T. Hinman, J. Zeisel, and S. Blackler, “The Escort system: a safety monitor for people living with Alzheimer's disease,” IEEE Perv. Comp., vol. 10, no. 2, 2011, pp. 68-77.

[39] L. Robinson, K. Brittain, S. Lindsay, D. Jackson, and P. Olivier, “Keeping In Touch Everyday (KITE) project: developing assistive technologies with people with dementia and their carers to promote independence,” Int. Psychogeriatrics, vol. 21, no. 3, 2009, pp. 494-502.

[40] Cornet, Victor P., and Richard J. Holden. "Systematic review of smartphone-based passive sensing for health and wellbeing." Journal of biomedical informatics 77 (2018): 120-132.

[41] Ernsting, Clemens, Stephan U. Dombrowski, et al., "Using smartphones and health apps to change and manage health behaviors: a population-based survey." Journal of medical Internet research 19, no. 4 (2017): e101.

[42] Boonstra, Tjeerd W., Jennifer Nicholas, et al., "Using mobile phone sensor technology for mental health research: integrated analysis to identify hidden challenges and potential solutions." Journal of medical Internet research 20, no. 7 (2018): e10131.

[43] Boukhechba, Mehdi, Philip Chow, Karl Fua, Bethany A. Teachman, and Laura E. Barnes. "Predicting social anxiety from global positioning system traces of college students: feasibility study." JMIR mental health 5, no. 3 (2018): e10101.

[44] Majumder, Sumit, and M. Jamal Deen. "Smartphone sensors for health monitoring and diagnosis." Sensors 19, no. 9 (2019): 2164.

[45] Higgins, John P. "Smartphone applications for patients' health and fitness." The American journal of medicine 129, no. 1 (2016): 11-19.

[46] R. Rawassizadeh, B. A. Price, and M. Petre, “Wearables: has the age of smartwatches finally arrived?” Communications of the ACM, vol. 58, no. 1, 2015, pp. 45-47.

[47] Foster, Kenneth R., and John Torous. "The opportunity and obstacles for smartwatches and wearable sensors." IEEE pulse 10, no. 1 (2019): 22-25.

[48] Silvera-Tawil, David, M. Sazzad Hussain, and Jane Li. "Emerging technologies for precision health: An insight into sensing technologies for health and wellbeing." Smart Health 15 (2020): 100100.

[49] Jacquemin, Charlotte, Hervé Servy, et al., "Physical activity assessment using an activity tracker in patients with rheumatoid arthritis and axial spondyloarthritis: prospective observational study." JMIR mHealth and uHealth 6, no. 1 (2018): e1.

[50] Leese, Jenny, Graham G. Macdonald, et al. "Using Physical Activity Trackers in Arthritis Self-Management: A Qualitative Study of Patient and Rehabilitation Professional Perspectives." Arthritis care & research 71, no. 2 (2019): 227-236.

[51] Gilmore, Sarah J., Andrew J. Hahne, Megan Davidson, and Jodie A. McClelland. "Physical activity patterns of patients immediately after lumbar surgery." Disability and Rehabilitation (2019): 1-7.

[52] Twiggs, Joshua, Lucy Salmon, et al., "Measurement of physical activity in the pre-and early post-operative period after total knee arthroplasty for Osteoarthritis using a Fitbit Flex device." Medical Engineering & Physics 51 (2018): 31-40.

[53] Deka, Pallav, Bunny Pozehl, Joseph F. Norman, and Deepak Khazanchi. "Feasibility of using the Fitbit® Charge HR in validating self-reported exercise diaries in a community setting in patients with heart failure." European Journal of Cardiovascular Nursing 17, no. 7 (2018): 605-611.

[54] Glance, David G., Esther Ooi, Ye'elah Berman, Charlotte F. Glance, and Hugh R. Barrett. "Impact of a digital activity tracker-based workplace activity program on health and wellbeing." In Proceedings of the 6th International Conference on Digital Health Conference, pp. 37-41. 2016.

[55] Z. Wang, et al., “A review of wearable technologies for elderly care that can accurately track indoor position, recognize physical activities and monitor vital signs in real time,” Sensors 17, no. 2, 2017, pp. 341.

[56] Galperin, Irina, Inbar Hillel, Silvia Del Din, Esther MJ Bekkers, Alice Nieuwboer, Giovanni Abbruzzese, Laura Avanzino et al. "Associations between daily-living physical activity and laboratory-based assessments of motor severity in patients with falls and Parkinson's disease." Parkinsonism & Related Disorders 62 (2019): 85-90.

[57] López-Blanco, Roberto, Miguel A. Velasco, Antonio Méndez-Guerrero, Juan Pablo Romero, María Dolores Del Castillo, J. Ignacio Serrano, Julián Benito-León, Félix Bermejo-Pareja, and Eduardo Rocon. "Essential tremor quantification based on the combined use of a smartphone and a smartwatch: The NetMD study." Journal of neuroscience methods 303 (2018): 95-102.

[58] Pulliam, Christopher L., Dustin A. Heldman, Elizabeth B. Brokaw, Thomas O. Mera, Zoltan K. Mari, and Michelle A. Burack. "Continuous assessment of levodopa response in Parkinson's disease using wearable motion sensors." IEEE Transactions on Biomedical Engineering 65, no. 1 (2017): 159-164.

[59] B. Reeder, and A. David, “Health at hand: a systematic review of smart watch uses for health and wellness,” Journal of Biomedical Informatics 63, 2016, pp. 269-276.

[60] de Arriba-Pérez, Francisco, Juan M. Santos-Gago, Manuel Caeiro-Rodríguez, and Mateo Ramos-Merino. "Study of stress detection and proposal of stress-related features using commercial-off-the-shelf wrist wearables." Journal of Ambient Intelligence and Humanized Computing 10, no. 12 (2019): 4925-4945.

[61] Shah, Yash, Jocelyn Dunn, Erich Huebner, and Steven Landry. "Wearables data integration: Data-driven modeling to adjust for differences in Jawbone and Fitbit estimations of steps, calories, and resting heart-rate." Computers in Industry 86 (2017): 72-81.

[62] de Zambotti, Massimiliano, Aimee Goldstone, Stephanie Claudatos, Ian M. Colrain, and Fiona C. Baker. "A validation study of Fitbit Charge 2™ compared with polysomnography in adults." Chronobiology international 35, no. 4 (2018): 465-476.

[63] Castner, Jessica, Manoj J. Mammen, Carla R. Jungquist, Olivia Licata, John J. Pender, Gregory E. Wilding, and Sanjay Sethi. "Validation of fitness tracker for sleep measures in women with asthma." Journal of Asthma 56, no. 7 (2019): 719-730.

[64] Winslow, Brent D., Nam Nguyen, and Kimberly E. Venta. "Improved mental acuity forecasting with an individualized quantitative sleep model." Frontiers in neurology 8 (2017): 160.

[65] Boukhechba, Mehdi, Lihua Cai, Congyu Wu, and Laura E. Barnes. "ActiPPG: using deep neural networks for activity recognition from wrist-worn photoplethysmography (PPG) sensors." Smart Health 14 (2019): 100082.

[66] J. Penders, M. Altini, C. Hoof, and E. Dy, “Wearable sensors for healthier pregnancies,” Proceedings of the IEEE 103, no. 2 (2015): 179-191.

[67] Grym, Kirsi, Hannakaisa Niela-Vilén, et al. "Feasibility of smart wristbands for continuous monitoring during pregnancy and one month after birth." BMC pregnancy and childbirth 19, no. 1 (2019): 1-9.

[68] Johansson, Dongni, Fredrik Ohlsson, David Krýsl, Bertil Rydenhag, Madeleine Czarnecki, Niclas Gustafsson, Jan Wipenmyr, Tomas McKelvey, and Kristina Malmgren. "Tonic-clonic seizure detection using accelerometry-based wearable sensors: A prospective, video-EEG controlled study." Seizure 65 (2019): 48-54.

[69] Cole, Casey A., Dien Anshari, Victoria Lambert, James F. Thrasher, and Homayoun Valafar. "Detecting smoking events using accelerometer data collected via smartwatch technology: Validation study." JMIR mHealth and uHealth 5, no. 12 (2017): e189.

[70] Orme, Mark W., Amie E. Weedon, et al. "Findings of the chronic obstructive pulmonary disease-sitting and exacerbations trial (COPD-SEAT) in reducing sedentary time using wearable and mobile technologies with educational support: Randomized controlled feasibility trial." JMIR mHealth and uHealth 6, no. 4 (2018): e84.

[71] Hooke, Mary C., Laura Gilchrist, et al. "Use of a fitness tracker to promote physical activity in children with acute lymphoblastic leukemia." Pediatric blood & cancer 63, no. 4 (2016): 684-689.

[72] Tsao, Liuxing, Liang Ma, and Christian-Thomas Papp. "Using non-invasive wearable sensors to estimate perceived fatigue level in manual material handling tasks." In International Conference on Applied Human Factors and Ergonomics, pp. 65-74. Springer, Cham, 2018.

[73] M. Swan, “Sensor mania! the internet of things, wearable computing, objective metrics, and the quantified self 2.0,” Journal of Sensor and Actuator Networks, vol. 1, no. 3, 2012, pp. 217-253.

[74] Baig, Mirza Mansoor, Shereen Afifi, Hamid GholamHosseini, and Farhaan Mirza. "A systematic review of wearable sensors and IoT-based monitoring applications for older adults–a focus on ageing population and independent living." Journal of medical systems 43, no. 8 (2019): 233.
[75] Niknejad, Naghmeh, Waidah Binti Ismail, Abbas Mardani, Huchang Liao, and Imran Ghani. "A comprehensive overview of smart wearables: The state of the art literature, recent advances, and future challenges." Engineering Applications of Artificial Intelligence 90 (2020): 103529.
[76] Tobore, Igbe, Jingzhen Li, Liu Yuhang, Yousef Al-Handarish, Abhishek Kandwal, Zedong Nie, and Lei Wang. "Deep learning intervention for health care challenges: some biomedical domain considerations." JMIR mHealth and uHealth 7, no. 8 (2019): e11966.
[77] N. Hammerla, S. Halloran, and T. Plötz, "Deep, convolutional, and recurrent models for human activity recognition using wearables," International Joint Conference on Artificial Intelligence, AAAI Press, 2016.
[78] F. Ordóñez and D. Roggen, "Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition", Sensors, vol. 16.1, pp. 115, 2016.
[79] Santos, Guto Leoni, Patricia Takako Endo, Kayo Henrique de Carvalho Monteiro, Elisson da Silva Rocha, Ivanovitch Silva, and Theo Lynn. "Accelerometer-based human fall detection using convolutional neural networks." Sensors 19, no. 7 (2019): 1644.
[80] Rad, Nastaran Mohammadian, Seyed Mostafa Kia, Calogero Zarbo, Twan van Laarhoven, Giuseppe Jurman, Paola Venuti, Elena Marchiori, and Cesare Furlanello. "Deep learning for automatic stereotypical motor movement detection using wearable sensors in autism spectrum disorders." Signal Processing 144 (2018): 180-191.
[81] Sathyanarayana, Aarti, Shafiq Joty, Luis Fernandez-Luque, Ferda Ofli, Jaideep Srivastava, Ahmed Elmagarmid, Teresa Arora, and Shahrad Taheri. "Sleep quality prediction from wearable data using deep learning." JMIR mHealth and uHealth 4, no. 4 (2016): e125.
[82] Shashikumar, Supreeth Prajwal, Amit J. Shah, Qiao Li, Gari D. Clifford, and Shamim Nemati. "A deep learning approach to monitoring and detecting atrial fibrillation using wearable technology." In 2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), pp. 141-144. IEEE, 2017.
[83] T. Kirste, A. Hoffmeyer, P. Koldrack, A. Bauer, S. Schubert, S. Schröder, and S. Teipel, "Detecting the effect of Alzheimer's disease on everyday motion behavior," Journal of Alzheimer's Disease, vol. 38, no. 1, pp. 121-132, 2014.
[84] Nweke, H.F., Teh, Y.W., Al-Garadi, M.A. and Alo, U.R., 2018. Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges. Expert Systems with Applications, 105, pp.233-261.
[85] R. Alam, J. Dugan, et al., "BESI: reliable and heterogeneous sensing and intervention for in-home health applications," IEEE/ACM Int Conf on Conn Health: Appl, Sys and Engr Tech (CHASE), IEEE, 2017.
[86] M. Ma, R. Alam, et al., "M^2G: A monitor of monitoring systems with ground truth validation features for research-oriented residential applications," Int Conf on Mobile Ad Hoc and Sensor Systems, IEEE, 2017.
[87] J. Cohen-Mansfield, "Conceptualization of agitation: results based on the Cohen-Mansfield agitation inventory and the agitation behavior mapping instrument," International Psychogeriatrics 8, no. S3, 1997, pp. 309-315.
[88] E. Koss, M. Weiner, C. Ernesto, J. Cohen-Mansfield, et al., "Assessing patterns of agitation in Alzheimer's disease patients with the Cohen-Mansfield Agitation Inventory. The Alzheimer's Disease Cooperative Study," Alzheimer Disease and Associated Disorders 11, 1997, pp. S45-50.
[89] M. Colombo, et al., "Behavioral and psychotic symptoms of dementia (BPSD) improvements in a special care unit: a factor analysis," Archives of Gerontology and Geriatrics, vol. 44, pp. 113-120, 2007.

[90] M. Carrillo, E. Dishman, and T. Plowman, “Everyday technologies for Alzheimer's disease care: Research findings, directions, and challenges,” Alzheimer's & Dementia, vol. 5, no. 6, pp. 479-488, 2009.
[91] A. Bharucha, et al., “Intelligent assistive technology applications to dementia care: current capabilities, limitations, and future challenges,” American Journal of Geriatric Psychiatry, vol. 17, no. 2, pp. 88-104, 2009.
[92] K. Kim, M. Hassan, S. Na, and E. Huh, “Dementia wandering detection and activity recognition algorithm using tri-axial accelerometer sensors,” Ubiquitous Info Tech & Applications (ICUT), pp. 1-5. IEEE, 2009.
[93] G. Nagels, S. Engelborghs, et al., “Actigraphic measurement of agitated behaviour in dementia,” International Journal of Geriatric Psychiatry, vol. 21, no. 4, pp. 388-393, 2006.
[94] A. Bankole, M. Anderson, et al., “Continuous, non-invasive assessment of agitation in dementia using inertial body sensors,” Wireless Health, p. 1, IEEE, 2011.
[95] R. Alam, J. Gong, et al., “Motion biomarkers for early detection of dementia-related agitation,” Proceedings of the Workshop on Digital Biomarkers, pp. 15-20. ACM, 2017.
[96] R. Alam, M. Anderson, A. Bankole, and J. Lach, “Inferring physical agitation in dementia using smartwatch and sequential behavior models,” IEEE EMBS Biomed. & Health Info. (BHI), pp. 170, 2018.
[97] R. Alam, A. Bankole, et al., “Multiple-instance learning for sparse behavior modeling from wearables: Toward dementia-related agitation prediction”, Int Conf of the IEEE Engr in Med and Bio Society (EMBC), 2019.
[98] A Bankole, M Anderson, N Homdee, R Alam, N Fyffe, H Goins, T Newbold, T Smith-Jackson, and J Lach, “BESI: Behavioral & environmental sensing & intervention for dementia caregiver empowerment - Phases 1 & 2,” SAGE American Journal of Alzheimer’s Disease & Other Dementias, 2020.
[99] N Homdee, R Alam, J Hayes, J Park, S Wolfe, J Lach, “Agitation monitoring and prevention system for dementia caregiver empower¬ment,” Computer, IEEE Computer Society, 2019.
[100] R Alam, N Homdee, S Wolfe, J Hayes, and J Lach, “BESI: Behavior learning and tracking with wearable and in-home sensors - A dementia case-study,” IEEE/ACM Internet-of-Things Design and Implementation, CPS-week 2019.
[101] M. Pijnenburg et al., “Monitoring asthma in children,” Euro. Resp. Jour., vol. 45, no. 4, pp. 906-925, 2015.
[102] R. Cowie et al., “The effect of a peak flow-based action plan in the prevention of exacerbations of asthma,” Chest, vol. 112, no. 6, 1997.
[103] S. Steinle et al., “Personal exposure monitoring of PM2.5 in indoor and outdoor microenvironments,” Sci. of Total Env., vol. 508, pp. 383, 2015.
[104] S. Steinle et al., “Quantifying human exposure to air pollution – moving from static monitoring to spatio-temporally resolved personal exposure assessment,” Sci. of Total Env, vol. 443, pp. 184-193, 2013.
[105] W. McDonnell et al., “Prediction of lung function response for populations exposed to a wide range of ozone conditions,” Inhalation Toxicology, vol. 24, no. 10, pp. 619-633, 2012.
[106] W. McDonnell et al., “Ozone-induced respiratory symptoms: exposure-response models and association with lung function,” Euro. Resp. Jour., vol. 14, no. 4, pp. 845-853, 1999.
[107] M. Miller et al., “Standardisation of spirometry,” Euro. Resp. J., vol. 26, no. 2, pp. 319-338, 2005.
[108] J. Dieffenderfer et al., “Low-power wearable systems for continuous monitoring of environment and health for chronic respiratory disease,” IEEE J. of Bio. and Health Info., vol. 20, no. 5, pp. 1251, 2016.
[109] Y. Khan et al., “Monitoring of vital signs with flexible and wearable medical devices,” Advanced Materials, vol. 28, no. 22, pp. 4373, 2016.
[110] F. Al-Khalidi et al., “Respiration rate monitoring methods: A review,” Pediatric Pulmonology, vol. 46, no. 6, pp. 523-529, 2011.

[111] C. Brüser et al., “Ambient and unobtrusive cardiorespiratory monitoring techniques,” IEEE Reviews in Biomed. Eng., vol. 8, pp. 30-43, 2015.
[112] C. Massaroni et al., “Contact-based methods for measuring respiratory rate,” Sensors, vol. 19, no. 4, pp. 908, 2019.
[113] J. Witt et al., “Measurement of exercise ventilation by a portable respiratory inductive plethysmograph,” Resp. Physiology & Neurobio., vol. 154, no. 3, pp. 389-395, 2006.
[114] L. Kent et al., “Validity and reliability of cardiorespiratory measurements recorded by the LifeShirt during exercise tests,” Respiratory Physiology & Neurobio., vol. 167, no. 2, pp. 162-167, 2009.
[115] R. Dumond et al., “Estimation of respiratory volume from thoracoabdominal breathing distances: Comparison of two models of machine learning,” Euro. J. of Applied Physio., vol. 117, no. 8, pp. 1533, 2017.
[116] C. Merritt, H. Nagle, and E. Grant, “Textile-based capacitive sensors for respiration monitoring,” IEEE Sensors Jour., vol. 9, no. 1, pp. 71, 2008.
[117] C. Voscopoulos et al., “Evaluation of a novel noninvasive respiration monitor providing continuous measurement of minute ventilation in ambulatory subjects in variety of clinical scenarios,” Anes. Anal., vol. 117, 2013.
[118] G. Moody et al., “Derivation of respiratory signals from multi-lead ECGs,” Computers in Cardiology, vol. 12, no. 1985, pp. 113-116, 1985.
[119] R. Bailón, L. Sornmo, and P. Laguna, “A robust method for ECG-based estimation of the respiratory frequency during stress testing,” IEEE Trans. on Biomed. Eng., vol. 53, no. 7, pp. 1273-1285, 2006.
[120] J. Boyle et al., “Automatic detection of respiration rate from ambulatory single-lead ECG,” IEEE Trans. Info. Tech. in Bio., vol.13, no. 6, 2009.
[121] D. Labate et al., “Empirical mode decomposition vs. wavelet decomposition for the extraction of respiratory signal from single-channel ECG: A comparison,” IEEE Sensors Jour., vol. 13, no. 7, pp. 2666, 2013.
[122] C. Orphanidou et al., “Data fusion for estimating respiratory rate from a single-lead ECG,” Biomed. Sig. Proc. and Control, vol. 8, no. 1, 2013.
[123] C. O’Brien and C. Heneghan, “A comparison of algorithms for estimation of a respiratory signal from the surface electrocardiogram,” Comp. in Bio. and Med., vol. 37, no. 3, pp. 305-314, 2007.
[124] E. Helfenbein et al., “Development of three methods for extracting respiration from the surface ECG: A review,” Jour. of Electrocardio., vol. 47, no. 6, pp. 819, 2014.
[125] P. Langley et al., “Principal component analysis as a tool for analyzing beat-to-beat changes in ECG features: application to ECG-derived respiration,” IEEE Trans. on Biomed. Eng., vol. 57, no. 4, pp. 821, 2009.
[126] D. Widjaja et al., “Application of kernel principal component analysis for single-lead-ECG-derived respiration,” IEEE Trans. on Biomed. Eng., vol. 59, no. 4, pp. 1169, 2012.
[127] P. Charlton et al., “An assessment of algorithms to estimate respiratory rate from the electrocardiogram and photoplethysmogram,” Physio. Measurements., vol. 37, no. 4, pp. 610, 2016.
[128] M. Pimentel et al., “Toward a robust estimation of respiratory rate from pulse oximeters,” IEEE Trans. on Biomed. Eng., vol. 64, no. 8, 2016.
[129] N. Lepine et al., “Robust respiration rate estimation using adaptive Kalman filtering with textile ECG sensor and accelerometer,” in Proc. of the Int. Conference of the IEEE EMBS (EMBC’16), pp. 3797, 2016.
[130] C. Rodes et al., “Predicting adult pulmonary ventilation volume and wearing compliance by on-board accelerometry during personal level exposure assessments,” Atmos. Environment, vol. 57, pp. 126, 2012.
[131] J. Kawahara et al., “Estimation of daily inhalation rate in preschool children using a tri-axial accelerometer: a pilot study,” Sci. of Total Env., vol. 409, no. 16, 2011.
[132] D. John, J. Staudenmayer, and P. Freedson, “Simple to complex modeling of breathing volume using a motion sensor,” Sci. of Total Env., vol. 454, pp. 184-188, 2013.

[133] R. Alam et al., “Non-invasive inference of minute ventilation using wearable ECG and Gaussian process regression,” in Proc. of the IEEE EMBS Int. Conf. on Bio. & Health Informatics (BHI’19), pp. 1-4, 2019.
[134] R. Alam et al., “Inferring respiratory minute volume from wrist motion,” in Proc. of the Int. Conf. of the IEEE EMBS (EMBC’19), pp. 6935, 2019.
[135] R Alam, D Peden, and J Lach, “Wearable respiration monitoring: Interpretable inference with context and sensor biomarkers,” IEEE Journal of Biomedical and Health Informatics (JBHI), [Under Review] 2020.
[136] R. Balani, “Energy consumption analysis for bluetooth, wifi and cellular networks,” Online: http://nesl.ee.ucla.edu/uploads/document/paperupload/254/PowerAnalysis.pdf, 2007.
[137] “libpebble2”, [Online], Available: https://github.com/pebble/libpebble2, [Accessed: 12-Dec-2016].
[138] “Distributed messaging - zeromq”, [Online], Available: http://zeromq.org/, [Accessed: 04-Oct-2016].
[139] M. Bertier, O. Marin, and P. Sens, “Implementation and performance evaluation of an adaptable failure detector,” International Conference on Dependable Systems and Networks, IEEE, 2002, pp. 354-363.
[140] “Supervisor: a process control system”, [Online], Available: http://supervisord.org/, [Accessed: 05-Nov-2016].
[141] J. Pan and W. Tompkins, A Real-Time QRS Detection Algorithm, IEEE Transactions on Biomedical Engineering, 32(3): 230-236, March 1985.
[142] L. Breiman, “Random forests,” Machine Learning, vol. 45, no. 1, 2001.
[143] C. Strobl et al., “Conditional variable importance for random forests,” BMC Bioinformatics, vol. 9, no. 1, pp. 307, 2008.
[144] R. Schapire and Y. Singer, “Improved boosting algorithms using confidence-rated predictions,” Machine Learning, vol. 37, no. 3, pp. 297, 1999.
[145] S. Hochreiter and J. Schmidhuber, “Long short-term memory”, Neural Comp, vol. 9.8, pp. 1735-1780, 1997.
[146] B. Frénay, and M. Verleysen, “Classification in the presence of label noise: a survey,” IEEE Transactions on Neural Networks and Learning Systems 25, no. 5, 2013, pp. 845-869.
[147] N. Natarajan, I. Dhillon, P. Ravikumar, and A. Tewari, “Learning with noisy labels,” NIPS, pp. 1196, 2013.
[148] J. Amores, “Multiple instance classification: Review, taxonomy and comparative study,” Artificial Intelligence, 201, pp. 81-105, 2013.
[149] S. Andrews, et al., “Support vector machines for multiple-instance learning,” NIPS, pp. 577-584, 2003.
[150] P. Viola, et al., “Multiple instances boosting for object detection,” NIPS, pp. 1417-1424, 2005.
[151] J. Nelder and R. Wedderburn, “Generalized linear models,” J. of Royal Statistical Society: Series A (General), vol. 135, no. 3, pp. 370, 1972.
[152] R. Tibshirani, “Regression shrinkage and selection via the lasso,” J. of Royal Stat. Soc.: Series B (Method.), vol. 58, no. 1, pp. 267-288, 1996.
[153] H. Zou and T. Hastie, “Regularization and variable selection via the elastic net,” J. of Royal Stat. Soc.: Series B, vol. 67, no. 2, pp. 301, 2005.
[154] A. Smola and B. Schölkopf, “A tutorial on support vector regression,” Statistics and Computing, vol. 14, no. 3, pp. 199-222, 2004.
[155] J. Quiñonero-Candela and C. Rasmussen, “A unifying view of sparse approximate Gaussian process regression,” J. of Machine Learning Research, vol. 6, no. Dec, pp. 1939-1959, 2005.
[156] M. Lázaro-Gredilla et al., “Sparse spectrum Gaussian process regression,” J. of Machine Learning Research, vol. 11, pp. 1865-1881, 2010.
[157] J. Goldberger et al., “Neighborhood components analysis,” Advances in Neural Info. Proc. Sys., pp. 513-520, 2005.
[158] W. Yang, K. Wang, and W. Zuo, “Neighborhood component feature selection for high-dimensional data,” J. of Comp., vol. 7, no. 1, 2012.
[159] M. Warmuth, J. Liao, and G. Rätsch, “Totally corrective boosting algorithms that maximize the margin,” in Proc. of ICML, pp. 1001-1008, 2006.

[160] Y. Yang, “Aggregating regression procedures to improve performance,” Bernoulli, vol.10, no. 1, pp. 25-47, 2004.

[161] J. Andreu-Perez et al., “Big data for health,” IEEE Jour. of Biomed. and Health Informatics, vol. 19, no. 4, pp. 1193-1208, 2015.

[162] M. Ribeiro, S. Singh, and C. Guestrin, ““Why should I trust you?” Explaining the predictions of any classifier,” in Proc. of the ACM SIGKDD Int. Conf. on Know. Disc. and Data Mining, pp. 1135, 2016.