

**COMBINING WEARABLES AND NEARABLES FOR PATIENT STATE
ANALYSIS**

A Dissertation
Presented to
The Academic Faculty

By

Pradyumna Byappanahalli Suresha

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in
Machine Learning

Georgia Institute of Technology

December 2022

COMBINING WEARABLES AND NEARABLES FOR PATIENT STATE ANALYSIS

Thesis committee:

Dr. Gari D. Clifford
The Wallace H. Coulter Department of
Biomedical Engineering
Georgia Institute of Technology

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

Dr. David V. Anderson
School of Electrical and Computer
Engineering
Georgia Institute of Technology

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

Dr. Ali Barhami Rad
Department of Biomedical Informatics,
School of Medicine
Emory University

Date approved: August 9, 2022

To work alone you have the right, and not to the fruits. Do not be impelled by the fruits of your work. May you not have any inclination for inaction.

Bhagawan Shri Krishna

To Amma, Anna and Pranathi

ACKNOWLEDGMENTS

This thesis would not be possible without the advice and help from numerous people who were part of my PhD journey. First and foremost, I would like to give my gratitude to my PhD advisor Prof. Gari D. Clifford, who has been my *Guru* for the last five years. He placed immense faith in me by taking me in as his PhD student and gave me much responsibility throughout the years which helped me improve as a researcher. He trusted me with hard problems and nudged me in the right direction when I got stuck in rabbit holes. He has been a great mentor and a person to know, and I will always consider him as my role model.

I would like to thank my PhD thesis defense committee members Dr. David V. Anderson, Dr. Omer T. Inan, Dr. Thomas Ploetz and Dr. Ali Bahrami Rad for reviewing my thesis dissertation and providing invaluable feedback regarding my work. Specifically, Dr. Anderson and Dr. Inan, as my PhD proposal committee and dissertation reading committee members helped shape my PhD thesis and the problem statement. For this, I am extremely grateful. I thank Dr. Ploetz for speaking to me about my research, for sharing his ideas on potential future work and providing crucial feedback. I thank Dr. Bahrami Rad for being a mentor in the last couple of years of my PhD at Emory University, for always asking thought provoking questions and providing feedback on my work. Furthermore, I am thankful to Dr. Faramarz Fekri for serving on my PhD preliminary examination committee and reviewing my examination report and presentations.

The Department of Biomedical Informatics (BMI) has been an integral part of my PhD journey and the colleagues at BMI have provided crucial support for my success. I thank Dr. Erik Reinertsen, Dr. Supreeth Prajwal Shashikumar, Dr. Samaneh Nasiri, Dr. Ayse Selin Cakmak, Dr. Camilo Valderrama, Dr. Nasim Katebi, Mohsen Salari, Nicolas Shu, Zifan Jiang, Samuel Waters and Chaitra Hegde who

during their days as PhD students at BMI have influenced me immensely inside and outside of work. A special thanks to Dr. Giulia Da Poian who as a Postdoctoral scientist in BMI closely collaborated with me in my initial years. I thank Dr. Qiao Li, Dr. Reza Sameni, Dr. Matthew Reyna, Dr. Yashar Kiarashi, Dr. Justus Schwabedal, Dr. Erick Perez Alday, Dr. Salman Seyedi and Dr. Hyeokhyen Kwon who always welcomed my questions and provided guidance at various times in my PhD journey. I thank Robert Tweedy and Jim Kinney for their assistance throughout the years as departmental computing support specialists at BMI. I thank Chad Robichaux for helping me ingest Emory's patient medical records data into my analyses. I thank Tony, Whitney, Chris and Daniel for assisting me in the engineering aspects of my projects. Lastly, I thank the numerous masters and undergraduate students who have worked with me and provided assistance on various projects at BMI.

I have been blessed to have worked with some wonderful scientists, clinicians and collaborators during my PhD journey. I would like to thank Dr. Paul Garcia for guiding me regarding the SedLine Root monitors and teaching me all that I know about sedation and anesthesia. I would like to thank Dr. Daniel Tarquinio for providing me the wearable database and gold standard labels for the Rett syndrome project and educating me about Rett syndrome. I thank Dr. Jana Von Hehn and Heather O'Leary for assisting me in developing the machine learning models for Rett syndrome severity estimation. I thank Dr. Donald Bliwise for teaching me about sleep-disorders, polysomnography and sleep scoring. I thank Dr. Amit Shah and Dr. Viola Vaccarino for providing valuable feedback on my manuscripts in the Emory Twins Study. I thank Dr. Theresa Hamlin for providing an opportunity to work with the great people at The Center for Discovery (TCFD), NY. I would like to thank Jenny Foster for providing me all the information I need about Autism spectrum disorder and the residents at TCFD. A special thanks to Conor Anderson for being the best host during my time at TCFD, taking care of all my needs, being a

friend during my multiple stays at the center and showing me the beautiful upstate New York.

Outside of my PhD, I made numerous friends during my time at Georgia Tech. They have in one way or the other had a big impact on my personal growth and have provided great support. I thank all my friends who have been part of my life in these years. I would like to specially mention Nikhil, Girish, Supreeth, Suraj, Abhiraj, Omkar, Ishaan, Tarun, Ashwin and Pradosh for sharing fun conversations at home and helping me relax outside work. I will cherish all the dinner conversations, late night walks and the festival celebrations we did together. Hindu YUVA has been an integral part of my PhD journey. I shall cherish all the weekly meetings, events, summer and winter hikes, and week-long camps we did together. The interactions with fellow YUVAs were always rejuvenating and provided different perspectives on various topics. I have indeed learnt much from these interactions. For this, I am forever indebted and shall seek out the friendship of YUVA volunteers wherever I go. I am also thankful to all the HSS families in Georgia who have welcomed me to their homes and have treated me as their extended family member. I am indebted to their hospitality which eased my transition to the life in the USA.

I am forever indebted to *Amma, Anna*, Pranathi and my extended family for their unconditional love and support during my PhD. I am confident that my parents' blessings have had a huge role to play in all my successes. I am thankful for their wisdom and for helping me get through the toughest of times. I thank my sister for her unwavering support and for the confidence she placed in my abilities. I would like to thank Kshama for her love and support in the final leg of my PhD. Finally, I prostrate to the all-powerful, all-knowing Brahman for everything.

Om Sarve Bhavantu Sukhinah Sarve Santu Niraamayaah |

Sarve Bhadraanni Pashyantu Maa Kashcid-Duhkha-Bhaag-Bhavet |

Om Shaantih Shaantih Shaantih ||

TABLE OF CONTENTS

Acknowledgments	v
List of Tables	xii
List of Figures	xiii
List of Acronyms	xvii
Summary	xx
Chapter 1: Introduction	1
1.1 Thesis Statement	2
1.2 Major Contributions	2
1.2.1 Journal articles	4
1.2.2 Conference articles	4
1.3 Scope and organization of the dissertation	5
Chapter 2: Background and related work	6
2.1 Wearables and nearables	6
2.2 Missing data and data imputation in signals from wearables and nearables	8
2.3 Biological mechanisms underlying Rett syndrome	9

2.4	Non-contact monitoring systems: A literature review	12
2.5	Quantifying body movements in OSA using nearables	14
2.6	Sleep-wake detection	15
Chapter 3: Severity of Rett syndrome from wearables		19
3.1	Introduction	19
3.2	Materials and methods	22
3.2.1	Data collection	22
3.2.2	Missing data	23
3.2.3	Feature extraction	26
3.2.4	Rett syndrome severity classification	29
3.2.5	Feature popularity	31
3.3	Results	32
3.3.1	Rett syndrome severity classification	32
3.3.2	Feature popularity	34
3.4	Discussion	36
3.5	Conclusion	40
Chapter 4: Patient state analysis from nearables		41
4.1	Introduction	41
4.2	Materials and methods	44
4.2.1	System architecture	44
4.2.2	Data fusion	51
4.3	Applications	52

4.3.1	Obstructive sleep apnea classification using a passive infrared sensor	53
4.3.2	Medical equipment alarm classification using audio	56
4.3.3	Geolocation of humans in a built environment via Bluetooth	59
4.3.4	Ambient light logging	61
4.3.5	Ambient temperature and humidity logging	62
4.4	Results	63
4.4.1	Obstructive sleep apnea classification using a passive infrared sensor	63
4.4.2	Medical equipment alarm classification using audio	64
4.4.3	Geolocation of humans in a built environment via Bluetooth	67
4.4.4	Ambient light logging	67
4.4.5	Ambient temperature and humidity logging	68
4.5	Discussion and Conclusion	68
Chapter 5: Combining wearables and nearables for sleep-wake detection		71
5.1	Introduction	71
5.2	Materials and methods	71
5.2.1	Dataset	71
5.2.2	Preprocessing and feature extraction	72
5.2.3	Experiments	76
5.3	Results	83
5.4	Discussion and Conclusion	84
Chapter 6: Conclusion		86

6.1	Future work	88
6.1.1	Severity of Rett syndrome from wearables	88
6.1.2	Patient state analysis from nearables	88
6.1.3	Combining wearables and nearables for sleep-wake detection	89
6.2	Final remarks	89
	References	90

LIST OF TABLES

4.1	Class labels including the Musical Notes in International Organization for Standardization / International Electrotechnical Commission (ISO/IEC) 60601 – 1 – 8 Alarm and their count [<i>E</i> stands for the <i>Empty</i> class and <i>T</i> stands for the <i>Transition</i> class]	56
4.2	The human tracking experiment	59
4.3	The ambient light tracking experiment	61
4.4	Classification performance when different feature combinations were used [<i>Acc</i> = Accuracy, \mathcal{F}_1 = F1 Score]. p^* : number of features needed in a given feature combination (with \mathcal{P}) to maximize \mathcal{F}_1	63
4.5	Note Classification in Medical Equipment Alarm	64
5.1	Performance of baseline sleep-wake detectors (Sleep-Acu = Sleep Accuracy, Wake-Acu = Wake Accuracy, κ = Kappa, SD = Standard deviation, Acc = Three-axis acceleration, PPG = Photoplethysmogram, TL = Transfer learned, CPD = Change point detection method)	83
5.2	Performance of proposed sleep-wake detectors (Sleep-Acu = Sleep Accuracy, Wake-Acu = Wake Accuracy, κ = Kappa, SD = Standard deviation, PPG = Photoplethysmogram, TL = Transfer learned, Acc = Three-axis acceleration	83

LIST OF FIGURES

1.1	Daily behavior and sleep patterns in health normals and patients suffering from neurological and sleep disorders differ and can be captured via continuous monitoring physiology and behaviors	2
1.2	The thesis overview illustrating the different project pipelines described in this dissertation.	3
2.1	The data collection setup for detecting body movements during sleep	14
3.1	The project pipeline. We utilized the BioStamp [®] nPoint biosensor wearables to collect simultaneous ECG and three-axis acceleration data from multiple locations on the chest. We performed data pre-processing and implemented multiple data imputation techniques to improve data quality. We trained L1-regularized logistic regression classifier models and tuned the model weights using the imputed data. Finally, we visualized model performance and computed feature popularity scores.	19
3.2	Stochastic surrogate data imputation technique fills gaps in heart rate (HR) and activity count signals by choosing a contiguous HR segment in the gap's neighborhood of the same length as the gap.(A) We illustrate an example HR signal with missing data. (B) The HR signal contains a 25-second-long gap that has been identified. Based on the boundaries and length of the gap, two folded normal distributions are constructed for stochastically choosing a 25-second-long HR segment. (C) We perform a coin toss experiment, and based on the outcome, we sample from one of the two folded normal distributions and accordingly select a contiguous 25-second-long HR segment and copy it over to the gap. Further, we add a small noise to this imputed signal (we do not show it in the figure for convenience). (D) The imputed HR signal with no gaps.	24

3.3 The feature combination of heart rate variability (HRV) metrics, multiscale transfer entropy, and multiscale network representation achieves the highest area under the receiver operating curve (AUC), equal to 0.92. (A) The receiver operating curves (ROCs) for the 15 different Rett severity classifiers are provided. Each classifier uses a different subset-combination of the four feature sets, namely: (1) HRV metrics (H), (2) Actigraphy metrics (A), (3) Multiscale transfer entropy features (T), (4) Multiscale network representation features (N). The combination of H+T+N performed the best with a leave one-patient out cross-validation pooled-AUC equal to 0.92. The individual ROCs corresponding to the individual classifiers are shown using a combination of line styles and colors. The figure legend shows the pooled-AUC values for each feature combination. (B) A depiction of the number of patient-visits for each of the 20 patients showing the low-severity and high-severity patient-visits in different colors. (C) The top 10 most popular features used by the best classifier (H+T+N) for Rett severity classification are shown here, along with the corresponding feature coefficients. The mean deceleration capacity ($\mu_{PRSA-DC}$) is the most popular feature with a feature popularity score of 1 followed by $\sigma_{\tau_{Act \rightarrow HR}}^{(8)}$ and $\sigma_{\tau_{Act \rightarrow DC}}^{(8)}$ with feature popularity scores 0.97 and 0.93, respectively. In the top-10 most popular features for the H+T+N feature combination, two features were HRV-metrics ($\mu_{PRSA-DC}$ and σ_{PNN50}^2), three features were MSNR-features ($\sigma_{LOOP4(3)}^2$, $\sigma_{LOOP4(5)}^2$ and $\mu_{TRACE(10)}$) and the rest five were MSTE-features. 30

3.4 Data imputation techniques combined with first and second-order statistics improve classification performance. (A) An illustration of the effects of novel signal quality index-based ECG data imputation. We show the improvements in the average signal quality index of the electrocardiogram records for all 32 patient-visits. (B) The deceleration capacity (DC) values between 10 PM and 10 AM for the low-severity and high-severity Rett syndrome groups. We show the DC values between the 75th and 25th percentile for low-severity and high-severity Rett patient-visits. (C) We depict all edges corresponding to 4-loops in the networks constructed from trivariate time series (HR, DC, and Activity count signal) for a high-severity (left-red-panel) and a low-severity (right-blue-panel) patient at the 3rd coarse-graining time scale. For each patient, we show two surrogate LOOP4 networks at the following values for LOOP4: (1) $\mu_{LOOP4(3)} - (1.96 \times \sigma_{LOOP4(3)})$, (2) $\mu_{LOOP4(3)} + (1.96 \times \sigma_{LOOP4(3)})$. From the image, it is clear that both the number of 4-loops and the variance of the number of 4-loops are smaller for the low-severity patient compared to the high-severity patient. 33

3.5	The top-five most popular features for each of the 15 classifiers. . . .	35
4.1	Edge computing and ambient data capture system. (PIR: Passive Infrared Sensor; USB: Universal Serial Bus)	42
4.2	Human geolocation via Bluetooth. The translucent background colors indicate the ground truth values, and the corresponding room number has been labeled. (A) The received signal strength indicator (RSSI) values as captured by each of the nine Raspberry Pis during the experiment are shown. It is measured in decibels with reference to one milliwatt (dBm). (B) The corresponding probability values of the subject being in Rooms 1, 2, or 3 during the experiment. The blue color depicts Room 1, the green color depicts Room 2, and the red color depicts Room 3.	65
4.3	Observing the effects of cloud cover, sunrise, sunset, lights-ON, lights-OFF and buildings around the data collection site on the ambient light in a built environment for two continuous days (48 hours). The solid-orange bars (■) depict the amount of ambient light sensed by the color sensor in Lumen per square meter (LUX). The solid-blue circles (●) depict the local cloud cover in percentage of local sky covered by clouds. The sunrise and sunset times are indicated with the rising and setting sun symbols, respectively, using the upward and downward arrows. The yellow and black bulbs specify the lights-ON and lights-OFF times, respectively. The skyscraper symbol indicates the time when the Sun goes behind a skyscraper and causes a shadow onto the location where ambient light was being tracked. Data upload is depicted by binary values and a cloud node.	65

4.4	Comparison of temperature and humidity values captured by the Raspberry Pi (RPi) integrated DHT22 sensor and a commercial sensor. The top plot corresponds to temperature in each subplot, and the bottom plot corresponds to humidity values. (A) The dashed blue lines indicate the processed temperature and humidity values captured by the RPi integrated DHT22 sensor. The solid red lines indicate the corresponding values captured by the commercial sensor. (B) The correlation plots between values captured from the DHT22 sensor and the commercial sensor. The solid-blue circles (●) indicate individual temperature and humidity tuples. The linear fits on the data and their deviations from the 45° line are depicted in the plot. (C) The Bland-Altman plot between the two measurements. The solid black lines indicate the mean difference between the two measurements. The dashed black lines indicate the +1.96 and the -1.96 standard deviation (SD) lines for the difference between the two measurements.	66
5.1	Baseline sleep wake detectors. (PPG: Photoplethysmogram; CRC: Cardiorespiratory Coupling; CNN: Convolutional Neural Network; SHHS: Sleep Heart Health Study; PRV: Pulse Rate Variability; P _{sleep} : Probability of sleep derived from the CNN model.)	78
5.2	Convolutional neural network structure of the pre-trained model [181]. The input is a 50 × 18 cardiorespiratory coupling plot and the output layer contains two neurons representing the probabilities of wake (class ●) and sleep (class ○)	79
5.3	Proposed sleep wake detectors. (PPG: Photoplethysmogram; CRC: Cardiorespiratory Coupling; CNN: Convolutional Neural Network; SHHS: Sleep Heart Health Study; PRV: Pulse Rate Variability; P _{sleep} : Probability of sleep derived from the CNN model.)	82

LIST OF ACRONYMS

ACT	Activity Count
AHI	Apnea-Hypopnea Index
AUC	Area Under the Receiver Operating Curve
BAN	Body Area Network
BHI	Biomedical & Health Informatics
CGI-S	Clinical Global Impression - Severity
CNN	Convolutional Neural Network
CPD	Change Point Detection
CRC	Cardiorespiratory Coupling
D-V	Darbellow-Vajda
DC	Deceleration Capacity
DSS	Difference Signal Statistics
DTFT	Discrete Time Fourier Transform
ECG	Electrocardiogram
EEG	Electroencephalogram
FDA	Food and Drug Administration
GPIO	General Purpose Input Output
HIPAA	Health Insurance Portability and Accountability Act
HR	Heart Rate
HRV	Heart Rate Variability
ICU	Intensive Care Unit
IEC	International Electrotechnical Commission

IMU Inertial Measurement Unit
IR Infrared
IRB Institutional Review Board
ISC Internet System Consortium
ISO International Organization for Standardization
LASSO Least Absolute Shrinkage and Selection Operator
LOOCV Leave One Out Cross-Validation
LUX Lumen Per Square Meter
MAC Media Access Control
MECP2 Methyl-CpG Binding Protein 2
MFCC Mel Frequency Cepstral Coefficients
ML Machine Learning
MSE Multiscale Entropy
MSNR Multiscale Network Representation
MSTE Multiscale Transfer Entropy
NN Beat-To-Beat Interval
NoIR No Infrared
OR Operating Room
OSA Obstructive Sleep Apnea
PAT Peripheral Arterial Signal
PCHIP Piecewise Cubic Hermite Interpolating Polynomial
PIR Passive Infrared
PLMD Periodic Limb Movement Disorder
PLMI Periodic Limb Movement iNdex
PLMR Proceedings of Machine Learning Research
POSIX Portable Operating System Interface Time
PPG Photoplethysmogram

PRSA Phase Rectified Signal Averaging
PRV Pulse Rate Variability
PSG Polysomnography
REM Rapid Eye Movement
RF Radio Frequency
RIAV Respiratory Induced Amplitude Variation
ROC Receiver Operating Curve
RPi Raspberry Pi
RSSI Received Signal Strength Indicator
SD Standard Deviation
SQI Signal Quality Index
STFT Short Time Fourier Transform
TE Transfer Entropy
TL Transfer Learned
UK United Kingdom
USB Universal Serial Bus
UTC Coordinated Universal Time

SUMMARY

Modern patient care and therapeutics for neurological and sleep-disorder populations rely on continuous and remote patient monitoring protocols [1, 2]. This moves the patient care from hospitals to homes and therapy centers which allow for long term monitoring at a lower cost. The improvements in sensor technologies have resulted in low-cost, portable and reliable patient monitoring systems which increasingly allow for remote and passive patient monitoring. These patient monitoring systems can be categorized as wearables (on-body) and nearables (off-body). While wearables capture localized physiological data such as pulse rate, wrist acceleration and brain signals, nearables record global passive data including body movements, ambient sound and environmental variables. Together, wearables and nearables provide a more comprehensive understanding of the patient state.

The processing of data captured from wearables and nearables have multiple challenges including handling missing data, time synchronization between sensors and data fusion techniques for multimodal analysis. The research described in this thesis aims to address these issues while working on data captured in the wild. We specifically focus on neurological (Rett syndrome) and sleep disorder (Obstructive sleep apnea) cohorts to test the effectiveness of our methods. In the first part of this dissertation, a wearable based disease severity estimator is developed for Rett syndrome. The proposed method utilizes heart rate and chest-wall acceleration data captured continuously using a wearable biosensor at the comfort of the patients' homes for long periods (at least two days). We provide data cleaning, data imputing and feature extraction methods for such signals and perform binary classification using a least absolute shrinkage and selection operator (LASSO) model. By using the features that capture the interaction between heart rate and body movements, this approach achieved a high classification performance

with the area under the receiver operating curve (AUC) equal to 0.92. In the second part of this dissertation, we develop an edge computing and ambient data capture system for clinical and home environments. Using the data captured by this system, we perform obstructive sleep apnea detection in a mixed-disorder elderly male cohort of 32 patients using a logistic regression classifier. Further, we perform the medical equipment alarm classification, a multi-class classification task, in silent and noisy speech settings and obtain high accuracy using an XGBoost classifier. Finally, we perform room-level geolocation of humans in a built environment using the Bluetooth signal strength data captured by our system. The last part of this dissertation focuses on methods to combine data from wearables and nearables to assess patient state. We consider the problem of sleep-wake detection in a mixed-disorder elderly male cohort using a wrist worn wearable and a video camera. We capture photoplethysmogram (PPG), wrist acceleration and video based body movements from 79 participants and derive features from each data modality. We leverage transfer learning to perform sleep-wake detection from PPG and a CatBoost classifier to perform multimodal sleep-wake detection. Among the unimodal models, video based body movement model performs better than the PPG and acceleration models in terms of κ . Combining all data modalities provides the best sleep-wake detection performance.

The work described in this dissertation provides foundational methods for capturing, combining and analyzing data captured using wearables and nearables. The methods presented can be extended for performing disease severity estimation and patient state analysis in other disease cohorts for obtaining better understanding of the patient physiology and behavior.

CHAPTER 1

INTRODUCTION

Over the years, sensor technologies have played a critical role in patient monitoring in clinical and home environments. Further, ambient patient monitoring using wearable (on-body) and nearable (off-body) sensors is becoming more prevalent, especially in the neurological and sleep disorder (*NeuroSleep*) populations [3, 4]. Traditional patient care which involves hospital visits, administering questionnaires, and in-hospital monitoring in intensive care units and operation theaters do not suffice for managing *NeuroSleep* cohorts. These disorders need long term monitoring (up to several months) and dynamic care (real-time decision making and support) that cannot be effectively provided by the traditional methods at low-cost. Further, as shown in Figure 1.1, the daily behaviors and sleep patterns alter in *NeuroSleep* cohorts compared to healthy normals. Moreover, each cohort presents a unique challenge for their continuous monitoring.

Wearables, also known as on-body sensors have the advantage of being able to collect high-resolution physiological signal data continuously and remotely. Among other signals, wearables can collect electrocardiogram (ECG), electroencephalogram (EEG) and three-axis acceleration. On the other hand, we have nearables that can passively record signals related to human behavior and ambient environment including human body movements, body pose, location, speech, acoustic noise such as patient monitoring alarms, ambient temperature, ambient humidity and ambient light intensity for long periods of time (at least several months). Thus, in this dissertation, we describe methods for patient monitoring in *NeuroSleep* cohorts by using data from wearables and nearables. There is much value to data captured by both wearables and nearables; they are complimentary and combining them could

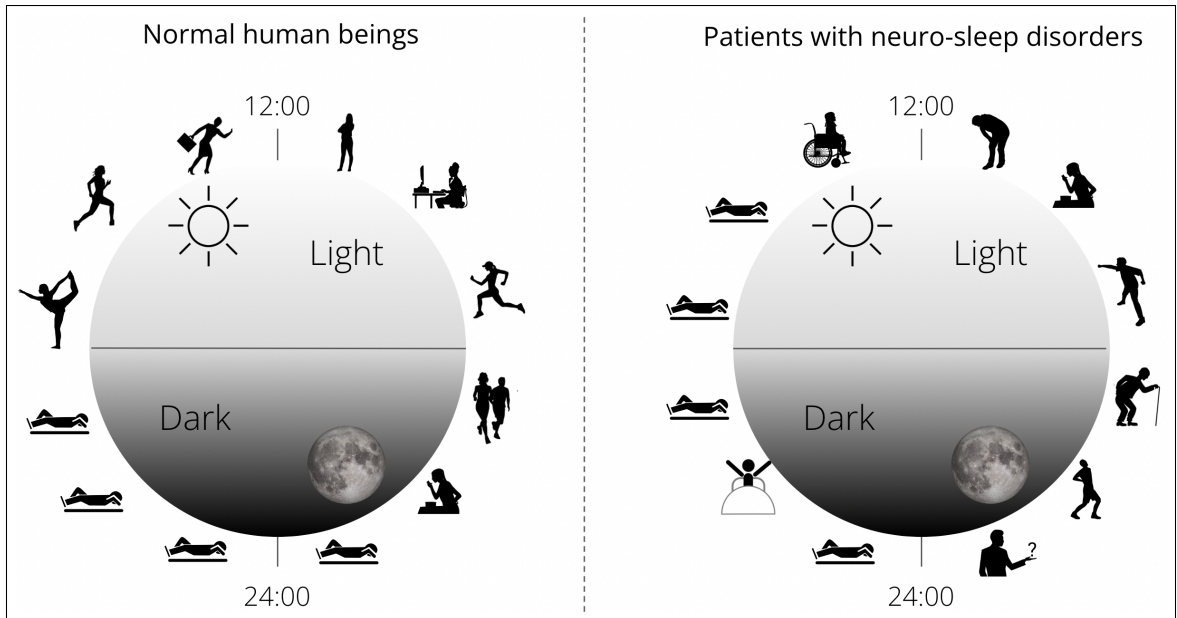


Figure 1.1 Daily behavior and sleep patterns in health normals and patients suffering from neurological and sleep disorders differ and can be captured via continuous monitoring physiology and behaviors

improve patient monitoring. Thus, we developed techniques to independently process the unimodal signals and to combine signals from wearables and nearables to perform multimodal patient monitoring.





1.1 Thesis Statement

The thesis statement is as follows:

“The combination of wearables and nearables allows for patient monitoring in diverse settings and could boost diagnostic power for patient state analysis. ”

1.2 Major Contributions

In this dissertation, we focused on three specific aims: (1) Wearable based disease severity estimation in Rett syndrome; (2) Generalized patient state analysis from nearables with applications in obstructive sleep apnea estimation; and (3) Combining wearable and nearable data for sleep-wake detection in a mixed-disorder

The disease or patient state being studied	Patient state variable to be estimated	Sensors	Physiology & behavior to be measured	Signals captured	Data cleaning & handling missing data	Machine Learning	Post-ML
Rett syndrome	Binarized CGI-S		HRV, actigraphy & interaction	ECG 3-axis Accel.	SQL-SS SSDI	LASSO	Biomarkers
Obstructive sleep apnea	Binarized AHI		Body movements	Motion timestamps	Body movement signal extraction	Logistic Regression	Feature Popularity
Passive ambient monitoring	Location & patient environment		Audio, T, H, LUX and RSSI	Audio, T, H, LUX and RSSI	Hand-crafted methods	XGBoost	GIF Visualization
Sleep-wake detection	Sleep-wake score		PRV, CRC, actigraphy & body movements	PPG Wrist-acceleration Video	Body movement signal extraction	CNN Transfer Learning CatBoost	Baseline model comparisons

 Wearables
 Nearables
 Wearables & Nearables

Figure 1.2 The thesis overview illustrating the different project pipelines described in this dissertation.

elderly male population.

- (1) In the first aim of the thesis, we developed methods to perform disease severity classification using a wearable. To demonstrate this, we performed Rett syndrome severity classification using the ECG and three-axis acceleration signals captured using the Biostamp®(MC10, Cambridge, MA, USA) nPoint on-body sensor [5].
- (2) In the second aim of the thesis, we built a Raspberry Pi-based hardware system to capture, collect and archive signals corresponding to human behavior and ambient environment. Further, we demonstrated the utility of this system via the following experiments - (1) Obstructive Sleep Apnea OSA classification using a passive infrared (PIR) sensor; (2) Medical equipment alarm note classification using audio; (3) Geolocation of humans in a built environment; (4) Ambient light logging; (5) Ambient temperature and humidity logging.
- (3) Lastly, in the third aim, we developed a multimodal signal processing and machine learning technique to combine the data captured using a wearable

(Empatica E4) and a wearable (video-polysomnography (PSG)) to perform sleep-wake detection.

The above contributions resulted in the following publications:

1.2.1 Journal articles

- **Suresha, P.B.**, Hegde, C., Jiang, Z. and Clifford, G.D., 2022. An Edge Computing and Ambient Data Capture System for Clinical and Home Environments. *Sensors*, 22(7), p.2511.
- **Suresha, P.B.**, Robichaux, C.J., Cassim, T.Z., García, P.S. and Clifford, G.D., 2022. Raspberry Pi-Based Data Archival System for Electroencephalogram Signals From the SedLine Root Device. *Anesthesia & Analgesia*, 134(2), pp.380-388.
- **Suresha, P.B.**, O’Leary, H., Tarquinio, D.C., Von Hehn, J. and Clifford, G.D., 2022. Rett syndrome severity estimation with the BioStamp nPoint using interactions between heart rate variability and body movement. *PLOS One*. [in review]
- **Suresha, P.B.**, Li, Q., Shah, A.J., Vaccarino, V., Bliwise, D. and Clifford, G.D., 2022, Transfer learning and multimodal data fusion of on and off-body sensors for sleep-wake detection. In *Machine Learning for Health (PLMR)* [in submission]

1.2.2 Conference articles

- **Suresha, P.B.**, Cakmak, A.S., Da Poian, G., Shah, A.J., Vaccarino, V., Bliwise, D. and Clifford, G.D., 2019, May. Obstructive Sleep Apnea Classification in a Mixed-Disorder Elderly Male Population Using a Low-Cost Off-Body Move-

ment Sensor. In 2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI) (pp. 1-4). IEEE.

In Figure 1.2 we present the various stages of all the thesis objectives. The nearables based patient state analysis methods are further divided into two sub-parts: (1) OSA detection and (2) Passive ambient monitoring.

1.3 Scope and organization of the dissertation

The dissertation is organized as follows. In chapter 2, we present relevant literature on wearables and nearables, the missing data problem, physiology in Rett syndrome, non-contact monitoring systems, quantifying body movements in OSA and sleep-wake detection. Chapter 3 focuses on the Rett syndrome severity classification problem, describing the data, methodology, results and conclusion. Further, it provides a detailed discussion on the measurement of physio-motor objective measures in Rett syndrome. Chapter 4 dives deep into the description of the edge-computing and ambient data capture system. Further, it presents five applications for data captured using nearables and includes the details regarding the OSA detection experiment. In chapter 5, we present the utility of video-based body movement features for sleep-wake detection in a mixed-disorder elderly male cohort. We use transfer learning to obtain improved sleep-wake detectors from wearables. Further, we provide methods to combine wearable and nearable data for improved sleep-wake detection. Finally, in Chapter 6, we provide concluding remarks and future directions.

CHAPTER 2

BACKGROUND AND RELATED WORK

The famous saying, "Standing on the shoulders of the giants", suggests that any scientific progress can only be made by using the understanding gained by previous works of numerous scientists. Thus, in this chapter we discuss the background for all the works presented in this dissertation. Further, we present related works that provide the context and motivation for the various works in this thesis.

2.1 Wearables and nearables

Since the turn of the 21st century, wearables (on-body sensors) have been at the forefront of non-traditional health monitoring systems [6, 7, 8]. These sensors collect high-resolution physiological signal data such as the ECG, EEG, body acceleration and galvanic skin response. Further, multiple wearables can be used in a network for remote patient monitoring. Body area networks (BANs) are one such system where multiple wearables continuously monitor human physiology and track the patient's health status [9]. BANs utilize wireless technologies including the ultra-wideband [10], Bluetooth [11], and Zigbee [12, 11] for this purpose. Wearables and BANs have numerous advantages to them:

- They capture high-resolution information regarding human physiology.
- They enable human physiology monitoring at the comfort of patient's homes.
- Wearables and BANs can act as a pre-screening tool for a more comprehensive diagnostic tool such as polysomnography [13].

They also suffer from the following drawbacks.

- Wearables perform localized measurements. For example, a wrist-worn accelerometer measures the acceleration of hand/wrist (local) and does not reliably measure overall body movement (global). Although a network of accelerometers will alleviate this problem, it comes at the cost of causing inconvenience to the patient as they have to wear multiple sensors on their body over long periods.
- Data from wearables are often corrupted by missing data due to motion artifacts [14] (the patient-induced noise in physiological signals by voluntary or involuntary bodily movements) and low compliance by the patients [15]. Human bodily movement causes motion artifacts in the physiological signals (say ECG) being captured by the wearable and thus leads to data degradation. Further, the wearer (a human) has to comply with a data acquisition schedule and follow the instructions diligently to generate *good quality* data.

Non-contact health monitoring systems or nearables, on the other hand, capture global signals (e.g., overall body movements via video camera [16]) and are less dependent on patient compliance for data capture. Further, due to non-dependency on patients or caregivers for data capture, these systems exhibit a higher compliance. Finally, nearables enable the monitoring of patient's ambient environment including temperature, humidity, lighting conditions and ambient sound. The patient's ambient environment could potentially act as stressors and have a negative impact on the patient's state and their healing process [17]. Thus, capturing ambient variables can add value to the patient state analysis pipeline. Wearables and nearables each have their own advantages and disadvantages. In this dissertation, we will develop methods to capture, process, analyze and combine the data captured using wearables and nearables for analyzing the patient's state.

2.2 Missing data and data imputation in signals from wearables and nearables

Wearable sensors have the advantage of extending the ability of assessing the patient's well-being outside the hospital environment and for a prolonged period, which might be longer than 24 hours (e.g., several days or months). One of the drawbacks of these technologies is that the patients are not 'under control' and the devices can be removed or switched off for a certain period during the evaluation. Device failures can also result in recordings with missing sections. Moreover, noise is a main issue when devices are used in an everyday life context, resulting in signals that might contain unanalyzable segments of different lengths. Note that the same missing data issues might also occur using more sophisticated ambulatory devices as well as when analyzing data coming from different datasets collected using different protocols and different recording lengths. In contrast, in the nearable sensors, the relative amount of missing data is generally lesser when compared to the wearable sensors. Since these sensors are not attached to patient's bodies, motion artifacts are not present in these data modalities. However, other types of noises such as shadow noise and periodic noise in video recordings, background noise in audio recordings (say when we are trying to record patient alarms) can occur in nearable sensor data resulting in missing data. Further, nearables that require an electrical connection will not record data when there is an electricity failure. Similar to wearables, device failures can occur in nearables as well resulting in recordings with missing sections. Another issue with nearables is that the captured data can correspond to multiple subjects or patients and this can often lead to confusion. In conclusion, employing signal quality checks and exploring data imputation techniques to improve data quality for patient state classification and prediction is crucial when working with data from wearables and nearables. In this thesis, we will develop two data imputation techniques for the ECG signal.

2.3 Biological mechanisms underlying Rett syndrome

Rett syndrome is a rare genetic neurological disorder that occurs primarily in girls [18]. Rett syndrome affects 1 in 10000 females causing psychomotor regression, stereotypical hand movements and dysautonomia [19]. Rett syndrome affects the neurophysiology of the patient and the presence of focal, multifocal, and generalized epileptiform abnormalities is characteristic of the EEG recorded in most individuals [20]. Further, individuals with Rett syndrome have significantly longer corrected QT intervals (QTc) and a higher incidence of prolonged QTc (> 0.450 milliseconds) than age-matched healthy females [21, 22, 23]. Finally, a pattern of breath holding (central or obstructive respiratory pauses) and hyperventilation (increased respiratory rate and effort) is frequently recorded in Rett syndrome.

A recent study by Singh et al. [24] showed that Rett patients are less adaptable to autonomic changes during the night which motivates our study - to discover physio-motor biomarkers discriminating high severity Rett patients from low severity Rett patients. Dysautonomia or autonomic dysfunction is the abnormal function of the autonomous nervous system. It adversely affects involuntary body functions, including blood-pumping by the heart, maintaining proper blood pressure and respiration. Unfortunately, dysautonomia is a cardinal feature in Rett syndrome [25, 26, 27, 28, 29, 30]. A principled approach to characterize dysautonomia is to utilize the ECG [26, 31, 32, 33]. Researchers have used the ECG to study the variations in breathing and heart rate (HR) in Rett girls [34, 35]. In our work, we captured the ECG signal from girls suffering from Rett syndrome using the BioStamp[®] nPoint wearable biosensors to capture the severity of autonomic dysfunction and, in turn, develop Rett severity classification models using ECG. Specifically, we used the heart rate variability (HRV) metrics for this purpose. The HRV is a physiological phenomenon of the variation of the time interval between the heartbeats. Typically,

it is measured using a set of metrics known as the HRV metrics. First, a peak picking algorithm is used to identify the R-peaks in the ECG signal and the time series known as the RR-interval signal is computed. The HRV metrics are then computed on this RR-interval signal. The HRV metrics include time domain, frequency domain and entropy based features. Dystonia is a movement disorder where a subset of muscles contract uncontrollably. The contractions cause the affected body parts to twist involuntarily, resulting in repetitive movements. In Rett syndrome, dystonia, psychomotor regression, and stereotypical hand movements are fundamental concerns and cause significant stress on patient and caregiver quality of life. We refer the readers to the following works [36, 37, 38, 39] for a detailed explanation of dystonia in Rett syndrome. The BioStamp[®] nPoint wearable biosensors measured the body movements of Rett patients by recording the three-axis acceleration signal via an accelerometer [40], an electromechanical sensor that senses static and dynamic forces of acceleration. To further characterize body movements, we derived rest activity metrics [41] and the cosinor rhythmometry features from the captured three-axis acceleration signals. Together, we call them actigraphy metrics. Apart from the above described unimodal features, we also investigated the usage of features derived from the interactions between HR and body movements. Finally, we trained machine learning models to perform Rett syndrome severity estimation utilizing all the above described features.

We posed the Rett syndrome severity estimation problem as a binary classification problem and classified low severity patient visits from high severity patient visits. The groundtruth labels were obtained by binarizing the physician assigned clinical global index - severity scale. The aggregated list of medications that were common to both groups included the following drugs: (1) Calcium carbonate, (2) Clonidine (Sedative and Antihypertensive drug), (3) Depakene (Anticonvulsant), (4) Diastat (Anxiolytic and Sedative), (5) Levocarnitine (treat carnitine deficiency),

(6) Miralax (Polyethylene glycol 3350), (7) Trileptal (Anticonvulsant), (8) Vimpat (anticonvulsant) and (9) Vitamin D. The high severity groups were further prescribed antihistamines, antibiotics, antioxidants, antifungals, calcium supplements, levothyroxine, escitalopram, magnesium oxide, esomeprazole and omeprazole to treat gastroesophageal reflux, budesonide, triamcinolone, multi-vitamins, vitamin B2 and other anticonvulsants and sedatives to treat seizures. We see that Rett syndrome patients are prescribed a large group of medications to control the various symptoms that the patients suffer. In our experiments in chapter 4 we set out to identify physiological biomarkers related to HRV in Rett syndrome. Hence, we review the effect of various medications listed above on patient HRV. The first major set of drugs prescribed for Rett patients include the anticonvulsants, antiepileptic and sedatives to treat seizure. Numerous studies [42, 43] have shown that the time domain HRV parameters are significantly suppressed in epilepsy patients groups with or without medications. Further, in some epileptic patients, antiepileptic drugs may help improve the cardiac autonomic function impairment [42]. Some Rett patients suffer from hypothyroidism for which thyroxine drugs have been prescribed. A study by Xing et al. [44] showed that the thyroxine drugs used to treat hypothyroidism partly improved autonomic function and did not suppress HRV. The gastroesophageal reflux in Rett syndrome is treated using the drugs esomeprazole and omeprazole and studies have shown that the drugs help improve autonomic function response on the cardiac activity and do not suppress HRV [45]. Triamcinolone is used to treat certain allergies and rheumatic disorder symptoms in Rett syndrome. A study by Cottin et al. [46] did not show a deleterious effect on autonomic cardiovascular control.

In chapters 4, we will further describe the Rett syndrome data in detail and discuss the machine learning approach to Rett syndrome severity estimation.

2.4 Non-contact monitoring systems: A literature review

Non-contact monitoring systems have been used to perform human body movement detection, vital sign measurement, ambient audio capturing, human geolocation and ambient environment monitoring. A popular way to perform non-contact monitoring of patients is to use the Doppler radar technology [47, 48]. The Doppler radar is a specialized radar system that can measure target displacement remotely by using the Doppler effect. It has been used for gait-assessment of older adults [49], capturing human respiration signal [50] and human vital sign measurement [51]. While it does an excellent job of detecting body movements and measuring the vital signs of a patient, it does not capture auditory cues or ambient environmental signatures.

Extensive research has been performed to study the effect of noise pollution on patient and staff health, the performance of staff, and patient safety in clinical environments [52, 53, 54, 55, 56]. However, very few works describe methods to capture privacy-preserving ambient sound in clinical and home environments. In particular, Guerlain et al. [57] presented a methodology for archiving multi-channel audio and video recordings of operating rooms (ORs) during surgeries to facilitate prospective studies of operative performance. To geolocate clinical personnel at fine-resolution, Azevedo-Coste et al. [58] proposed using multiple cameras installed in ORs along with a wireless network of inertial sensors. On the other hand, recently, there has been a surge in the development of radio frequency (RF)-based non-contact human movement detectors and geolocators [59, 60, 61].

Aschoff et al. [62] provided a comprehensive explanation about the effect of ambient temperature, humidity, and light intensity on the human circadian rhythm and sleep. There is a growing evidence [63, 64, 65, 66, 67, 68] about the effects of ambient environment on patient health. While few systems [69, 70] provided

methodology for capturing these ambient variables, we did not find works that captured concurrent patient motion, auditory noise and environmental signals.

The geolocation of humans has been performed using Bluetooth beacons and inertial measurement unit (IMU) sensors. Sato et al. [71] performed geolocation on six participants using the range-only extended Kalman filter simultaneous localization and mapping technique. Martín et al. [72] utilized the IMU sensor along with the Bluetooth beacon to obtain a room level (six-rooms) human localization while geolocating four participants. There exist multiple works that analyzed generic ambient environmental sounds including in-home sounds, emergency sirens and everyday sounds. We have described three specific studies that were related to our work on alarm note classification. Jain et al. [73] developed two prototype systems for in-home sound analysis and deployed each system in four different homes to recognize 23 in-home sounds and three outdoor sounds. Cantarini et al. [74] applied the harmonic percussive source separation technique to classify emergency siren sounds from road noise sounds. Wyatt et al. [75] deployed a BERT-based environmental sound classification model on an RPi Zero to identify six different everyday sounds (Knock, Laugh, Keyboard Typing, Cough, Keys Jangling and Snap). All three studies mentioned above described standalone audio capture and recognition systems for varied environmental settings and did not describe a multi-modal data acquisition system. There are few works in the literature that perform ambient environment data capture. Rashed et al. [70] described a medical platform for remote healthcare and assisted living. They utilized a DHT11 temperature and humidity sensor and a TSL235R light to frequency converter to keep track of the patient's environmental conditions.

In Chapter 3, we describe our non-contact monitoring system that captures patient privacy preserving signatures corresponding to body movement, ambient sound, illuminance, temperature and humidity. Further, we utilize this system

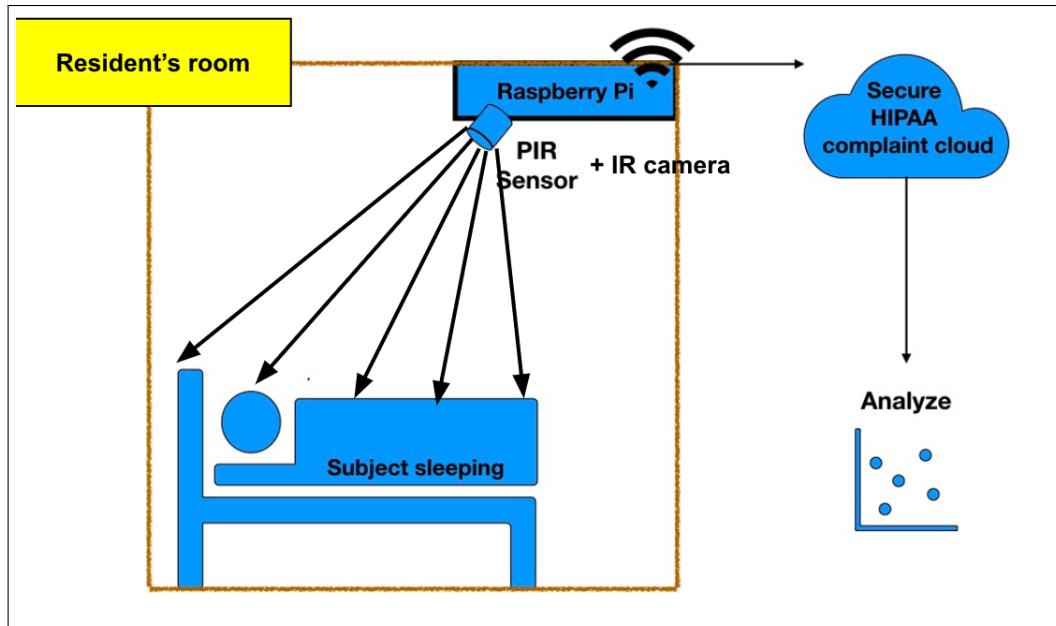


Figure 2.1 The data collection setup for detecting body movements during sleep for capturing passive ambient data from a cohort suffering from Autism spectrum disorder during their sleep for predicting their behavior during the day.

2.5 Quantifying body movements in OSA using nearables

OSA, a serious sleep disorder that causes people to temporarily stop breathing several times during the night, is typically diagnosed using PSG in a sleep-lab and more recently using portable sleep monitors [76]. OSA has a prevalence of 3 – 7% among the adult male population and the risk of having OSA increases with age [77].

While PSG still remains the standard test for the diagnosis of OSA and clinical sleep assessment, at-home monitoring using technologies similar to the ones specified in [78, 79, 80] have enormous potential for tracking disease progression, treatment compliance, response to therapy, and ongoing screening for critical workers (e.g. truck drivers, pilots etc). The adverse effects of OSA and its multiple possible complications have resulted in a substantial and growing demand by pa-

tients to access diagnostic studies and effective treatment [81]. Moreover, since OSA is a chronic disorder, sufferers often require lifelong monitoring and care [82]. Thus, in this work we propose the use of low-cost and easy-to-use (high compliance) sensors that can be adopted for follow-up monitoring as well as to help identify patients with OSA who would otherwise be undiagnosed and untreated.

The work by Roebuck et al. [83] provides a comprehensive report on the different data modalities one can use for OSA detection. Various unimodal and multimodal data driven approaches have been used in the past for OSA detection. For instance, while some researchers have used wearables to capture ECG in conjunction to oxygen saturation to train machine learning models to detect OSA [84, 85], others have relied on nearables including video monitoring, audio signals, mattress sensors and smartphones [16, 86, 87, 88, 89]. In chapter 4, we will introduce a new data modality that can be used in OSA detection. We will specifically introduce the PIR sensor and illustrate the system we develop to capture human body movement using this sensor. Further, we train a machine learning algorithm to classify patients with OSA from normals using the data collected with the PIR sensor.

2.6 Sleep-wake detection

Besides enabling complete sleep analysis, sleep-wake detection methods have numerous secondary applications. Two important applications include: (1) The measurement of sleep disorder severity. For instance, the apnea hypopnea index (disease severity metric for OSA) computation involves the measurement of *total sleep time* which can be derived from sleep-wake scores. (2) The quantification of sleep disturbances. Sleep disturbances (body movements and awakenings) have an impact on daily behaviors in neurological disease populations such as Autism spectrum disorder [90, 91]. These disturbances can be quantified using a sleep-wake detection method.

The clinical standard for sleep disorder diagnosis is a sleep study conducted at a sleep lab via an overnight PSG. It includes numerous EEG electrodes to monitor the brain signals, as well as electromyograms, electrooculograms, finger tip pulse oximetry, respiratory bands across the chest wall and lower abdomen, actigraphy, audio and video. On the other hand, as discussed earlier, sleep disorders such as OSA and periodic limb movement disorder (PLMD) which are diagnosed with the help of such sleep studies, depend on the accurate estimation of *total sleep time* for a reliable measurement of disease severity metrics. Further, many patients get admitted to the sleep labs to understand their sleep cycles better and in turn improve their sleep. Thus, accurate sleep-wake detection and the more general five-level (wake, rapid eye movement (REM) sleep, non-REM stage-1, non-REM stage-2 and non-REM stage-3) sleep-staging become highly important. The cumbersome and intrusive nature of the gold standard PSG procedure may affect the normal sleep of the admitted patients resulting in a less-reliable sleep study and disease severity estimation. Moreover, the ground truth sleep staging is typically performed by a human annotator (albeit experienced) and can result in noisy labels.

Over the last decade, data from wearables [92, 93, 94] including wrist-worn accelerometers and photoplethysmograms (PPGs) have provided promising results for sleep-wake detection. Further, nearables such as microphones [95], video cameras [96] and radar sensors [97, 98] have been utilized to reliably perform sleep-wake detection. Furthermore, considerable advancement has been made in the commercial space for sleep-wake detection and sleep disorder detection. Consequently, numerous hardware and software systems have received Food and Drug Administration (FDA) clearance for the above stated applications. WatchPAT[®] [99] is an FDA cleared home sleep apnea device that measures seven channels (peripheral arterial signal PAT[®], heart rate, oximetry, actigraphy, body position, snoring, and chest motion) for sleep apnea detection. Also, there exist sleep tracking mats

such as Withings [100] and Beddit [101] which track sleep, heart rate and breathing disturbances utilizing mats that go underneath the sleeping humans. In November 2021, The United Kingdom (UK) based medical device company Acurable obtained FDA clearance as a Class II medical device for their AcuPebble SA100, a small wearable device that enables automated, remote home testing to detect OSA in adults [102]. Another product to obtain FDA clearance for OSA screening is the DROWZLE[®] sleep apnea pre-screening device developed by the company Resonea. DROWZLE[®] consists of stand-alone smartphone-based software that records and analyzes respiratory patterns during sleep for the purpose of in-home screening for obstructive sleep apnea [103]. ResApp Health's SleepCheckRx app received FDA clearance as a prescription-only software-as-a-medical device [104] for at-home sleep test that screens adults for the risk of moderate to severe OSA by analyzing breathing and snore sounds recorded via an iPhone. While the FDA cleared systems discussed here illustrate a promising future for sleep apnea screening, there are other related key areas that need more work. Developing an accurate sleep-wake detector which is highly accurate (E.g. Cohen's kappa > 0.7), both in the general and the sleep-disorder suffering populations is still extremely valuable. Please refer subsection 5.2.3 for the description of the inter-rater reliability metric – Cohen's kappa [105]. The measurement of many other sleep disorders including the PLMD benefit from the metrics derived from a robust sleep-wake detector. Further, neurological conditions such as Autism and Rett syndrome greatly benefit from sleep-wake detection as it allows for the measurement of night time circadian rhythm. Specifically, in Autism, alterations in circadian sleep-wake rhythm, reduced total sleep, longer sleep latency and nocturnal awakenings are observed [106, 107, 106, 108]. Moreover, these deviations from the normality have a negative impact on their day-time behaviors. A low quality sleep in the night can cause a higher incidence of aggressive behavior [109]. Furthermore, the behavioral issues in Autism pose a

challenge for patient's sleep monitoring. For instance, a wearable watch designed to capture PPG and three-axis acceleration can be damaged by the study participant due to the behavioral issues. Thus, there is a need to explore passive patient monitoring methods for sleep analysis. Further, the sleep related metrics could potentially help understand and predict daily behavior density for these patients. Toward this, in Chapter 5, we explore the usage of the passive patient-privacy preserving sleep monitoring system using a video-camera. Further, we describe a multimodal algorithm, that combines data captured using a wrist-worn wearable and the video-camera to perform improved sleep-wake detection.

CHAPTER 3

SEVERITY OF RETT SYNDROME FROM WEARABLES

3.1 Introduction

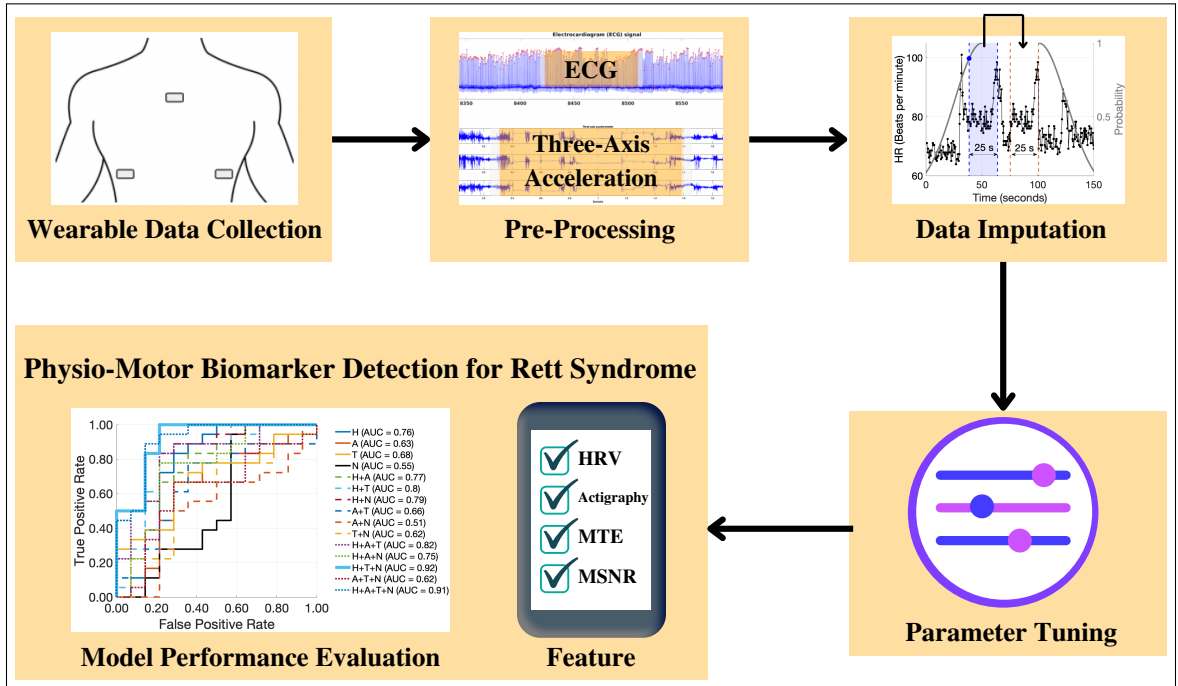


Figure 3.1 The project pipeline. We utilized the BioStamp[®] nPoint biosensor wearables to collect simultaneous ECG and three-axis acceleration data from multiple locations on the chest. We performed data pre-processing and implemented multiple data imputation techniques to improve data quality. We trained L1-regularized logistic regression classifier models and tuned the model weights using the imputed data. Finally, we visualized model performance and computed feature popularity scores.

In this chapter we will discuss methods to utilize data collected by wearables for patient state analysis. Specifically, we developed methods to estimate disease severity in Rett syndrome using signals captured using the MC10 Biostamp nPoint

wearable biosensor. The necessary background on the disease: Rett syndrome, its symptoms and its effect on patient's autonomic nervous system and motion were provided in section 2.3. To summarize, Rett syndrome is a disorder where more than 95% of cases are caused by mutations in the gene encoding the methyl-CpG binding protein 2 (MECP2), a transcriptional regulator involved in chromatin remodeling and the modulation of RNA splicing [110]. The disease is characterized by a period of apparently normal postnatal development followed by developmental delay and loss of acquired skills resulting in psychomotor regression, development of stereotypical hand movements, and dysautonomia [111]. It leads to the deterioration of the autonomous nervous system, impacting breathing regularity, HR, gut motility, and impairs motor planning and locomotion, resulting in significantly impaired mobility, no purposeful hand use, and largely absent verbal communication. There is no permanent cure for Rett syndrome in humans, and symptom management remains the standard of care [112]. When new drugs are discovered to alleviate specific Rett symptoms, clinical trials are conducted to learn about their efficacy, safety, and side effects. An essential step in measuring the efficacy of a drug or treatment method is to assess the associated benefits and risks through clinical trials. However, objective measures of symptom severity are not yet available for Rett syndrome or neurological conditions generally, and efforts to develop objective measures of autonomic symptoms could significantly enhance the ability to understand therapeutic efficacy. The key benefit we would like to see in Rett patients is improving their autonomous nervous system's function and locomotion. Amongst various indices measured in Rett clinical trials, the Clinical Global Impression - Severity (CGI-S) is used to measure overall disease severity in Rett subjects [113, 114]. The CGI-S is a 7-point Likert rating scale that reflects experts' clinical judgment of the patient based on the clinician's total experience with the Rett syndrome population. The CGI-S ranges from 1 to 7 and each score

corresponds to the following patient states: (1) normal, not at all ill, (2) borderline ill, (3) mildly ill, (4) moderately ill, (5) markedly ill, (6) severely ill, (7) amongst the most extremely ill. It has been widely used as an outcome measure in Rett syndrome and other neurodevelopmental disorders such as Autism and Fragile X Syndrome [113]. In our work, we measured CGI-S in all our patients during all clinic visits included in our experiments to assess their global clinical state.

Apart from the unimodal features described in section 2.3, interactions between the HR and body movement were quantified using multiscale transfer entropy (MSTE) [115, 116] and multiscale network representation (MSNR) [117]. The MSTE metrics measured the information flow between two simultaneously sampled time series at multiple time scales. In MSNR, we constructed network representations of simultaneously sampled three-dimensional time series at multiple time scales and derived network characteristics at each time scale. These network representations revealed more nuanced characteristics of the time series being analyzed.

The goal of this study was two-fold. First, we wanted to develop machine learning (ML) classification models to classify patients with low-severity Rett syndrome ($\text{CGI-S} \leq 4$) from patients with high-severity Rett syndrome ($\text{CGI-S} > 4$) based on the objective measures attained from a wearable biosensor. Second, through the classification experiment, we wanted the trained models to provide us with important features (physio-motor biomarkers) that could help us distinguish the two groups. Hence, we developed Rett syndrome severity classifier models based on raw data recordings using metrics derived from the following feature sets: (1) HRV metrics, (2) Actigraphy metrics, (3) MSTE-features, and (4) MSNR-features. We used the least absolute shrinkage and selection operator (LASSO) for model training and developed logistic regression models with the L1-penalty. We developed separate models for each of the four feature sets listed above and for all possible two, three, and four combinations of these feature sets. Thus, we developed 15 binary-

classification models for Rett syndrome severity classification. Finally, we listed the models' features that were important for Rett syndrome severity classification. We illustrate the complete pipeline of our work in Figure 3.1.

3.2 Materials and methods

3.2.1 Data collection

The dataset for this work was sourced from two Institutional Review Board (IRB)-approved studies: (1) The Triheptanoin-clinical trial [118] (2) The Outcome measures and biomarkers development study [119]. The data were collected between January 2016 and December 2018 - a three-year period. We used the body-worn patch BioStamp[®] (MC10 Inc., Cambridge, MA, USA) [120] to record ECG and three-axis acceleration from all the participants. While some ECG records were captured at a sampling rate of 125Hz, others were captured at a sampling rate of 250Hz. Concurrently, the three-axis acceleration records were captured at the sampling rates of 31.25Hz and 62.5Hz, respectively. These differences did not meaningfully influence the HRV and activity metrics we extracted [121]. We captured the ECG signal and the three-axis acceleration from the following four locations on the body: (1) Medial chest, (2) Left Hypochondrium, (3) Right Hypochondrium, and (4) Left Pectoralis. Per the protocols, all four locations were not used for all the participants, and only a subset of these locations was used for each participant. In conjunction to the signal data obtained from the biosensors, caretaker and physician surveys were conducted to obtain symptom severity for all 20 patients enrolled in the study. Specifically, the CGI-S scores were obtained through physician surveys to assign a binary label (low-severity vs. high-severity) for each patient-visit. A patient-visit was assigned to the low-severity category if the CGI-S ≤ 4 and was assigned to the high-severity category if the CGI-S > 4 . For each patient-visit we needed two consecutive days of signal data for the feature extraction. By applying this filter, we

obtained a total of 32 patient-visits with two consecutive days of signal data and the associated CGI-S label. Among the 32 patient-visits, we had 18 high-severity visits corresponding to 10 unique patients and 14 low-severity visits corresponding to 11 unique patients. One patient had both low-severity and high-severity visits. We considered each patient-visit a data point and thus had 32 data points with an associated binary label for model development and analysis.

Study approval

This study was approved by the Emory IRB (IRB00088492 : Outcome Measures and Biomarkers Development for Rett Syndrome and Multisite development of standardized assessments for use in clinical trials). A written informed consent was received prior to the participation from the parents of the patients.

3.2.2 Missing data

Considerable amounts of missing data were present in the dataset due to the following reasons: (1) Device charging and data upload, (2) Motion artifacts, and (3) In some cases, low compliance by the caretakers. Thus, we implemented three signal imputation techniques to improve data quality and increase the amount of available data for analysis. Namely,

- 1 Signal quality index-based ECG data imputation.
- 2 Data imputation for activity counts.
- 3 Stochastic surrogate data imputation.

Signal quality index-based ECG data imputation

The ECG signal recorded using the wearable patches contained sections of data corrupted by motion artifacts. To improve data quality, we sourced data from

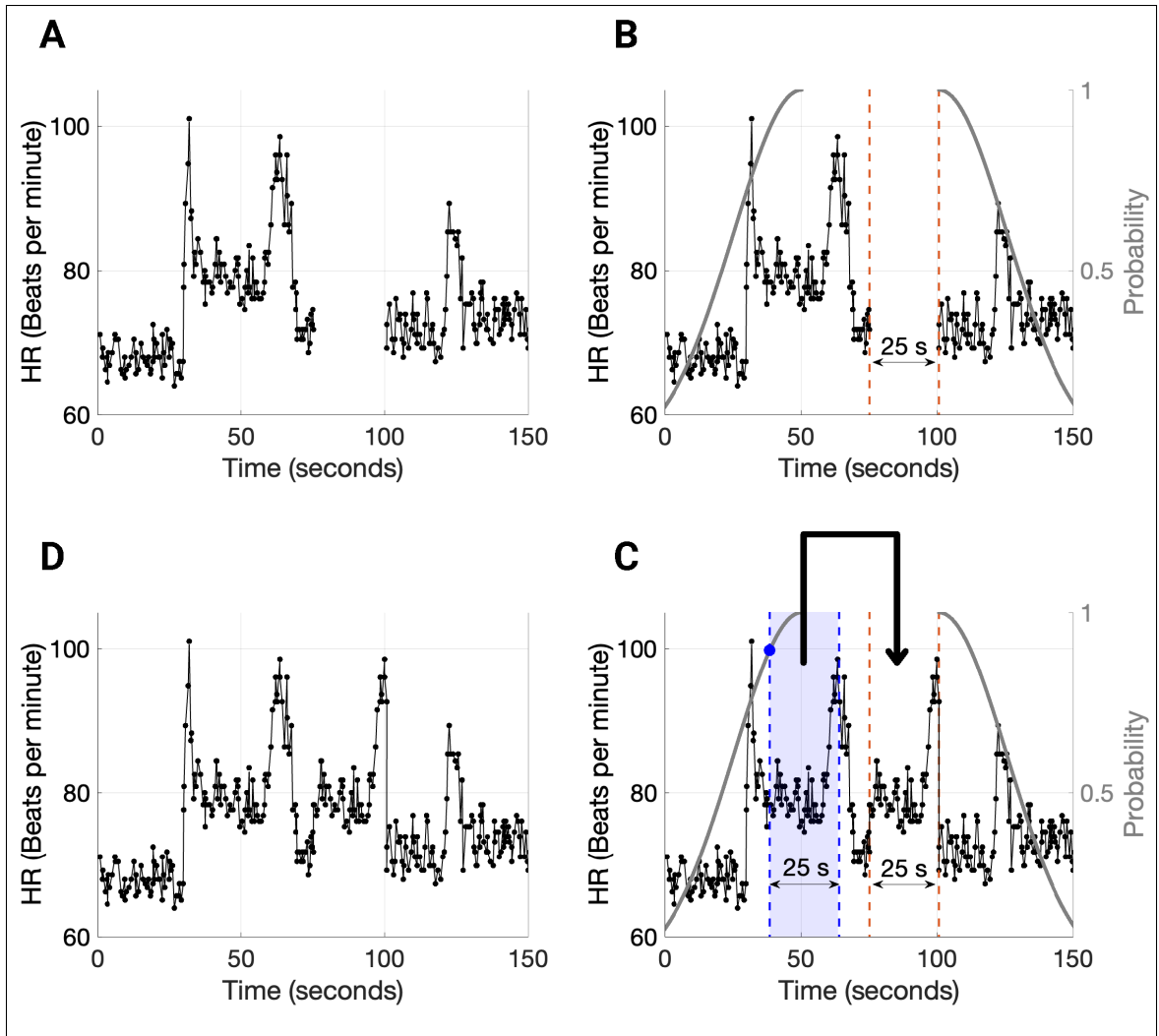


Figure 3.2 Stochastic surrogate data imputation technique fills gaps in heart rate (HR) and activity count signals by choosing a contiguous HR segment in the gap's neighborhood of the same length as the gap. (A) We illustrate an example HR signal with missing data. (B) The HR signal contains a 25-second-long gap that has been identified. Based on the boundaries and length of the gap, two folded normal distributions are constructed for stochastically choosing a 25-second-long HR segment. (C) We perform a coin toss experiment, and based on the outcome, we sample from one of the two folded normal distributions and accordingly select a contiguous 25-second-long HR segment and copy it over to the gap. Further, we add a small noise to this imputed signal (we do not show it in the figure for convenience). (D) The imputed HR signal with no gaps.

multiple sensors. As discussed earlier, we simultaneously captured ECG from up to four unique locations on the body. Thus, for a time-window t , say we had N ($N \leq 4$) ECG signal snippets recorded from N locations, we chose the one signal snippet with the highest Signal Quality Index (SQI). The SQI for ECG provides the percentage of beats that match when detected by multiple annotation generators with highly differing noise responses [122]. We refer the readers to Li et. al [123] for a detailed explanation of the SQI computation algorithm. If this value was greater than 0.75, we used it in our analysis; otherwise, we discarded all ECG data for the time-window t . By switching between N signals for each time-window t to form a single 1-D ECG signal, we maximized the amount of good data available for the analysis.

Data imputation for activity counts

When analyzing activity count signal in isolation, as per the scripts provided in the actigraphy toolbox [124], we combined the data to obtain one value per hour. When there was no data in a given hour, we imputed those samples using the Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) interpolation.

Stochastic surrogate data imputation

The computation of transfer entropy and multiscale network representation required us to impute the missing data in the 48-hour HRV features and the activity count signals. For this, we developed the surrogate data imputation method, a stochastic technique developed to impute missing data in a timeseries using data in the vicinity of the missing sections (or *gaps*). The data imputation algorithm works as follows. Given a time series (\mathcal{S}) and its timestamps (t), we find all the N gaps $g[i] \forall i \in \{1, 2, 3, \dots, N\}$ in \mathcal{S} which are greater than a fixed threshold th_g . The gaps are then sorted in increasing order. We impute the gaps with surrogate signal

snippets in increasing order of the gap length as follows. We denote the gap length in seconds as g_l , while t_b and t_e denote the time stamp where the gap begins and ends, respectively. Next, for each gap $g[i]$ we draw a sample x_r from the normal distribution with mean 0 and variance equal to 1, i.e., $x_r \sim \mathcal{N}(0, 1)$. The normal distribution from which we picked a sample is then mapped to the timeseries \mathcal{S} in the following way. We map the left half of the distribution ($0.5 \times$ folded normal distribution) to \mathcal{S} where $t < t_b - g_l$ and the right half of the distribution ($0.5 \times$ folded normal distribution) to \mathcal{S} where $t > t_e$. We illustrate this mapping of the Gaussian distribution onto the timeseries in Figure 3.2. Accordingly, we copy the signal snippet of length g_l starting from the point in time that corresponds to x_r on the timestamp signal t . We insert this copied signal in the gap $g[i]$ and add a noise signal which is 5% of the sample sampled from a Gaussian distribution with mean ($\mu_{\mathcal{S}}$) and variance ($\sigma_{\mathcal{S}}^2$) equal to the mean and variance of the signal \mathcal{S} . Further, we update both the timeseries \mathcal{S} and the corresponding timestamps t . This procedure is repeated iteratively until all gaps $\{g[i]\}$ in \mathcal{S} greater than th_g are imputed.

3.2.3 Feature extraction

We extracted HRV metrics from the ECG signals and actigraphy features from the three-axis accelerometer signals in the dataset. The HRV metrics were extracted using the open-source PhysioNet Cardiovascular Signal Toolbox provided by Vest et al. [122]. We extracted 24 distinct HRV metrics, including time-domain measures, frequency-domain measures, entropy measures, phase rectified signal averaging (PRSA) measures, and other non-linear metrics. We used the default window length settings provided in the toolbox and thus used a 300-second-long feature extraction window with a 30-second shift. We used the SQI based ECG data imputation to maximize the amount of good ECG data. For a given patient-visit, we computed the mean and variance of each HRV metric between the times 10 PM and 10 AM.

We chose this period to include ECG data during sleep and discard the noisier signal recorded during daytime and evenings. Thus, each HRV metric provided two features resulting in 48 features from 24 HRV metrics.

We extracted the actigraphy features from the z-axis of the acceleration signal from the right hypochondrium using the open-source Actigraphy Toolbox [124]. We converted the acceleration signal to activity counts using the Oakley method described by Borazio et al. [125]. First, we used Oakley’s method for converting accelerometer signals to activity count. We then extracted the following eight features using the toolbox: (1) Interday stability, (2) Intraday variability, (3) Least active 5 hours, (4) Most active 10 hours, (5) Rest activity, (6) Mesor, (7) Amplitude, and (8) Acrophase. The last three features were based on Cosinor Rhythmometry. The Actigraphy features needed two consecutive 24-hour periods (midnight to midnight) of data for feature computation. Thus, we identified the best two consecutive 24-hour periods with the least missing data for each patient-visit. If both days did not have at least 12-hours of acceleration data per day, those patient-visits were discarded. To impute missing data, we used the PCHIP interpolation.

Finally, we computed MSTE and MSNR features using 2-day consecutive HR and activity count signals. We utilized the *Stochastic surrogate data imputation* technique to impute missing data. We processed the HR signal, deceleration capacity (DC) of the RR-interval [126] signal, and the activity count (ACT) signal for computing the features. Transfer entropy (TE) from X to Y depicted as $TE_{X \rightarrow Y}$ is a measure of directional coupling between two concurrently sampled timeseries $X = \{x_1, x_2, \dots, x_N\}$ and $Y = \{y_1, y_2, \dots, y_N\}$. Formally, $TE_{X \rightarrow Y}$ is a reduction in uncertainty, given by the conditional entropy of y_i given its past values minus the conditional entropy of y_i given both its past values and past values of the other variable $y_{i-w}^{(l)}$:

$$TE_{X \rightarrow Y} = H(y_i | y_{i-w}^{(l)}) - H(y_i | y_{i-w}^{(l)}, x_{i-t}^{(k)}) \quad (3.1)$$

where i indicates a given point in time, t and w are the time lags in X and Y respectively, and k and l are the block lengths of past values in X and Y respectively. The k and l were both set to 1 to improve computational speed, and t and w were both set to 1 under the assumption that the maximum auto-transfer of information occurs from the data point in X immediately before the target value in Y , and vice versa. These choices of $k = l = t = w = 1$ are appropriate in biomedical experiments as the absolute values of auto-correlation functions tend to decrease monotonically as time lag increases [116]. In our experiments, we computed the TE between the following signals: (1) HR-DC, (2) HR-ACT, (3) DC-ACT, (4) DC-HR, (5) ACT-HR, and (6) ACT-DC. We computed these TE values at scales 1 to 10 using the coarse-graining algorithm [127] to obtain Multiscale Transfer Entropy (MSTE) features. The probability densities for the estimation of MSTE were estimated using the Darbellay-Vajda (D-V) adaptive partitioning algorithm [116, 117, 128]. Further, we computed 3D D-V partitioning using the HR-DC-ACT signals and computed MSNR features. The network representation features included the following 11 metrics: (1) Number of nodes (total number of nodes in the network), (2) Average degree (the average value of the degree of all nodes in the network, where the degree of a node is defined as the total number of its neighboring edges), (3) Number of loops (the total number of independent loops in the network, also known as the “cyclomatic number” or the number of edges that need to be removed so that the network cannot have cycles), (4) LOOP3 (the total number of loops of size 3 in the network), (5) LOOP4 (the total number of loops of size 4 in the network), (6) Average clustering coefficient - 1 (number of LOOP3s divided by the number of connected triples in the network), (7) Average clustering coefficient - 2 (the clustering coefficient $c(u)$ for node u can be defined as the ratio of the number of actual edges between the neighbors of u to the number of possible edges between them, and the average clustering coefficient $C(G)$ of a network is the average of $c(u)$

taken over all the nodes in the network), (8) Graph radius (the eccentricity of a node u is defined as $e(u) = \max\{d(u, v) : v \in V\}$, where the distance $d(u, v)$ is the length of the shortest path from u to v , and V is the set of all nodes; the graph radius is the minimum of eccentricities over all nodes in the network), (9) Spectral radius (the largest magnitude eigenvalue of the adjacency matrix of the network), (10) Trace (sum of the eigenvalues of the adjacency matrix, i.e., $\sum \lambda$), and (11) Energy (squared sum of the eigenvalues of the adjacency matrix A . More formally, the energy of a network G is: $E(G) = \sum_i^n \lambda_i^2$). We computed the above-described MSNR features at scales 1 to 10 using the coarse-graining technique [127]. Using the surrogate data imputation technique, we obtained 100 imputations for each patient-visit. Thus, we generated 3200 datapoints. We computed the MSTE and MSNR features for all the 3200 datapoints and then computed the mean and variance over 100 imputations for each patient-visit. In the end, we obtained 32 vectors of length 120 (60 mean values and 60 variance values) as the MSTE features and obtained 32 vectors of length 220 (110 mean values and 110 variance values) as the MSNR features. Thus, we obtained 48 HRV features, 8 Actigraphy features, 120 MSTE features, and 220 MSNR features for each of the 32 patient visits.

3.2.4 Rett syndrome severity classification

We developed separate models for each feature sub-group and combinations of feature sub-groups to obtain 15 classifiers corresponding to the following feature combinations: (1) HRV, (2) Actigraphy, (3) MSTE, (4) MSNR, (5) HRV + Actigraphy, (6) HRV + MSTE, (7) HRV + MSNR, (8) Actigraphy + MSTE, (9) Actigraphy + MSNR, (10) MSTE + MSNR, (11) HRV + Actigraphy + MSTE, (12) HRV + Actigraphy + MSNR, (13) HRV + MSTE + MSNR, (14) Actigraphy + MSTE + MSNR, and (15) HRV + Actigraphy + MSTE + MSNR. We used the LASSO based logistic regression classifier to assess the performance of different feature combinations.

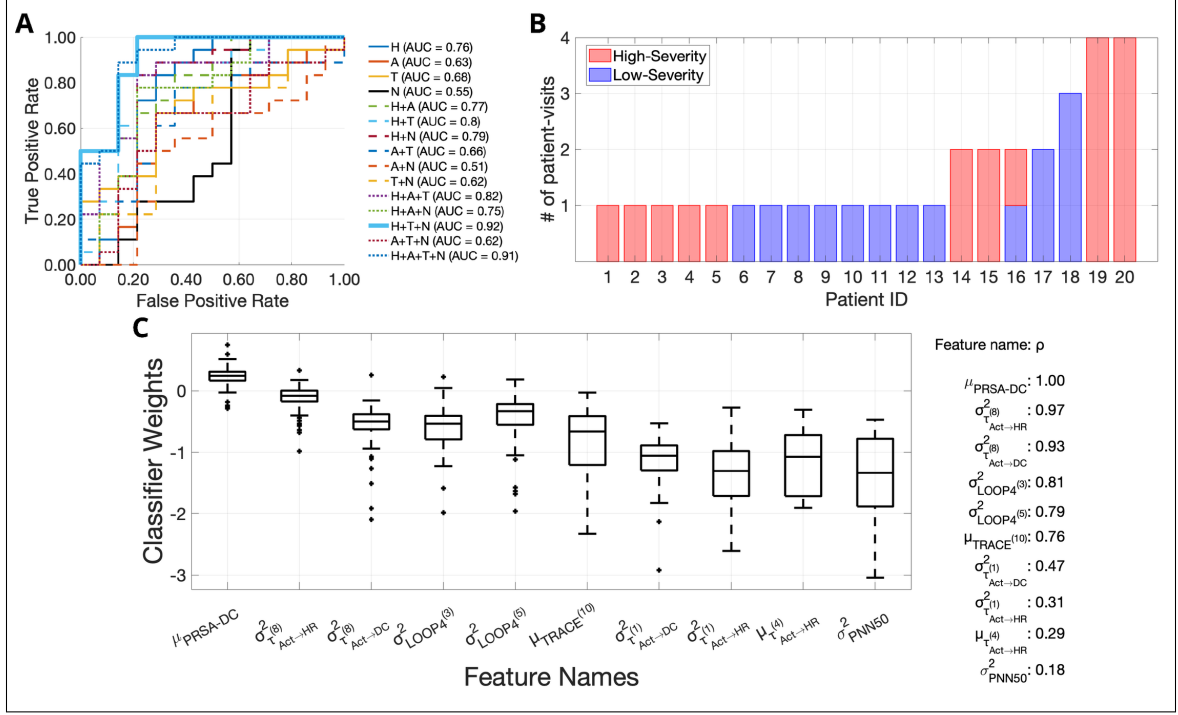


Figure 3.3 The feature combination of heart rate variability (HRV) metrics, multi-scale transfer entropy, and multiscale network representation achieves the highest area under the receiver operating curve (AUC), equal to 0.92. (A) The receiver operating curves (ROCs) for the 15 different Rett severity classifiers are provided. Each classifier uses a different subset-combination of the four feature sets, namely: (1) HRV metrics (H), (2) Actigraphy metrics (A), (3) Multiscale transfer entropy features (T), (4) Multiscale network representation features (N). The combination of H+T+N performed the best with a leave one-patient out cross-validation pooled-AUC equal to 0.92. The individual ROCs corresponding to the individual classifiers are shown using a combination of line styles and colors. The figure legend shows the pooled-AUC values for each feature combination. (B) A depiction of the number of patient-visits for each of the 20 patients showing the low-severity and high-severity patient-visits in different colors. (C) The top 10 most popular features used by the best classifier (H+T+N) for Rett severity classification are shown here, along with the corresponding feature coefficients. The mean deceleration capacity ($\mu_{PRSA-DC}$) is the most popular feature with a feature popularity score of 1 followed by $\sigma_{\tau_{Act \rightarrow HR}}^{(8)}$ and $\sigma_{\tau_{Act \rightarrow DC}}^{(8)}$ with feature popularity scores 0.97 and 0.93, respectively. In the top-10 most popular features for the H+T+N feature combination, two features were HRV-metrics ($\mu_{PRSA-DC}$ and σ_{PNN50}^2), three features were MSNR-features ($\sigma_{LOOP4^{(3)}}^2$, $\sigma_{LOOP4^{(5)}}^2$ and $\mu_{TRACE^{(10)}}$) and the rest five were MSTE-features.

The models were assessed via a leave-one-patient out cross-validation experiment. The hyperparameter tuning was performed using a within-training-three-fold cross-validation. We measured the classification performance using the metric - area under the receiver operating curve (AUC). Using this metric, we compared the 15 different classifiers and assessed their classification performance. For each feature combination, the entire classification experiment was repeated five times with different random seeds to account for variability in model coefficients that arose during hyperparameter tuning (the within-training-three-fold data split changed each time). The model's outputs for the five repetitions of the experiment were combined by computing the median value for the classification probability output. The final AUC for each feature combination was determined by comparing this median output against the ground truth labels.

3.2.5 Feature popularity

Apart from measuring classification performance, we computed a novel feature popularity score for the features used in classification, which allowed us to measure feature importance and compare features. Since all classification experiments comprised 20 patients, the leave-one-patient out cross-validation approach produced 20 models. As described previously we repeated the classification experiment five times for a given feature combination, resulting in $5 \times 20 = 100$ classifier models per feature combination. In these 100 models, for each feature f , we counted the number of models in which it had a non-zero LASSO coefficient. Then, the feature popularity score (ρ) for the feature f was computed as:

$$\rho = \frac{\# \text{ of nonzero coefficients for } f}{100} \quad (3.2)$$

This metric measured the popularity of the feature; the greater number of times the feature was picked by the LASSO-based classifiers, the more popular the feature was.

3.3 Results

The HRV, breathing, and physical activity are thought to influence Rett syndrome severity, but the underlying correlations are yet to be measured. Our experiments investigated the effect of HRV-metrics, actigraphy, MSTE, and MSNR features on Rett syndrome severity by assessing 15 binary classification models (Figure 3.3A) described in the section *Rett syndrome severity classification*. For this, as described in the section *Data collection*, we utilized a cohort of 20 patients with Rett syndrome, of which 10 patients had low-severity patient-visits, nine patients had high-severity patient-visits and one individual had both low-severity and high-severity patient-visits. We illustrate this in Figure 3.3B.

3.3.1 Rett syndrome severity classification

The best binary Rett severity classifier used the feature combination of HRV-metrics, MSTE-features, and MSNR-features, and obtained a pooled-AUC equal to 0.92. When we used the four feature sub-groups separately for classification, HRV-metrics performed the best with a pooled-AUC equal to 0.76. This was followed by MSTE-features (AUC = 0.68), Actigraphy-metrics (AUC = 0.63) and MSNR-features (AUC = 0.55) respectively. When we used 2-combinations of feature sets, the feature combination of HRV and MSTE performed the best with an AUC = 0.80. This was followed by the following 2-combinations of feature sets: (1) HRV + MSNR (AUC = 0.79), (2) HRV + Actigraphy (AUC = 0.77), (3) Actigraphy + MSTE (AUC = 0.66), (4) MSTE + MSNR (AUC = 0.62), and (5) Actigraphy + MSNR (AUC = 0.51). When we used 3-combinations of feature sets, we obtained the following

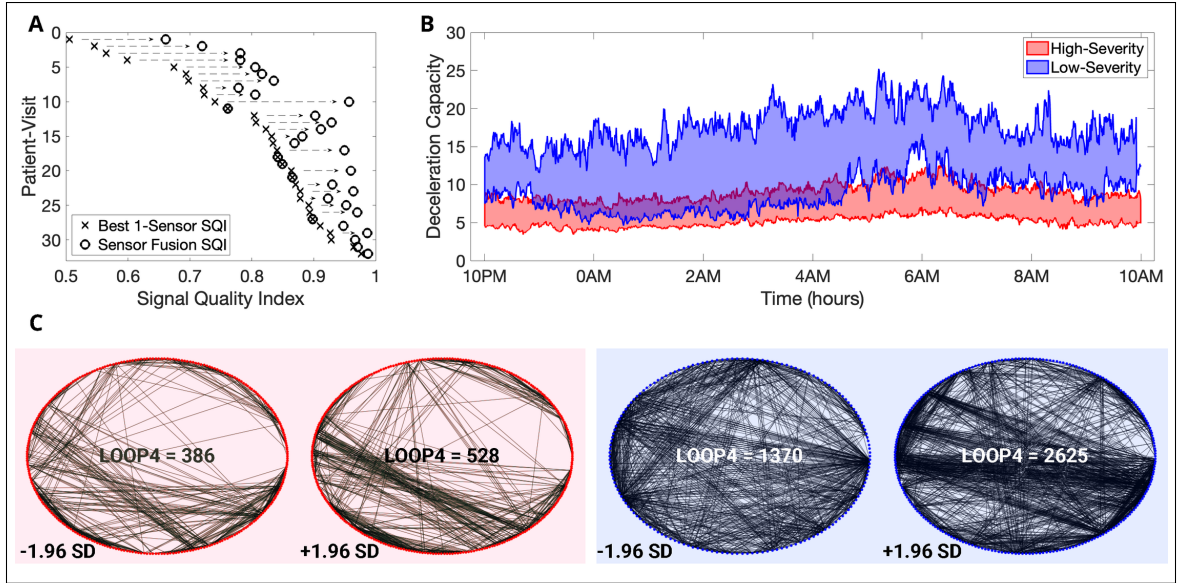


Figure 3.4 Data imputation techniques combined with first and second-order statistics improve classification performance. (A) An illustration of the effects of novel signal quality index-based ECG data imputation. We show the improvements in the average signal quality index of the electrocardiogram records for all 32 patient-visits. (B) The deceleration capacity (DC) values between 10 PM and 10 AM for the low-severity and high-severity Rett syndrome groups. We show the DC values between the 75th and 25th percentile for low-severity and high-severity Rett patient-visits. (C) We depict all edges corresponding to 4-loops in the networks constructed from trivariate time series (HR, DC, and Activity count signal) for a high-severity (left-red-panel) and a low-severity (right-blue-panel) patient at the 3rd coarse-graining time scale. For each patient, we show two surrogate LOOP4 networks at the following values for LOOP4: (1) $\mu_{LOOP4^{(3)}} - (1.96 \times \sigma_{LOOP4^{(3)}})$, (2) $\mu_{LOOP4^{(3)}} + (1.96 \times \sigma_{LOOP4^{(3)}})$. From the image, it is clear that both the number of 4-loops and the variance of the number of 4-loops are smaller for the low-severity patient compared to the high-severity patient.

descending order of classification performances: (1) HRV + MSTE + MSNR (AUC = 0.92), (2) HRV + Actigraphy + MSTE (AUC = 0.82), (3) HRV + Actigraphy + MSNR (AUC = 0.75), (4) Actigraphy + MSTE + MSNR (AUC = 0.62). Finally, when all feature subsets were used together, we obtained an AUC = 0.91, only second to the combination of HRV-metrics, MSTE-features and MTNR-features (AUC = 0.92).

3.3.2 Feature popularity

The mean deceleration capacity ($\mu_{PRSA-DC}$) was the most popular feature for Rett syndrome severity classification. Using the novel formula described in the section *Feature popularity*, we extracted the top-10 most popular features utilized by the best classifier (HRV, MSTE, and MSNR) and the distribution of their corresponding weights (Figure 3.3C). The feature $\mu_{PRSA-DC}$ came out on top with a feature popularity score $\rho = 1.00$. It was followed by the following 9 features: (1) Variance (across all surrogate representations) of transfer entropy from the activity counts signal to the HR signal at the 8th coarse-graining scale ($\sigma_{\tau_{Act \rightarrow HR}^{(8)}}^2$) ($\rho = 0.97$), (2) Variance (across all surrogate representations) of transfer entropy from the activity counts signal to the deceleration capacity signal at the 8th coarse-graining scale ($\sigma_{\tau_{Act \rightarrow DC}^{(8)}}^2$) ($\rho = 0.93$), (3) Variance (across all surrogate representations) of the number of 4-loops in the network representation of the tuple - (HR signal, activity count signal, DC signal) at the 3rd coarse-graining scale ($\sigma_{LOOP4(3)}^2$) ($\rho = 0.81$), (4) Variance (across all surrogate representations) of the number of 4-loops in the network representation of the tuple - (HR signal, activity count signal, DC signal) at the 5th coarse-graining scale ($\sigma_{LOOP4(5)}^2$) ($\rho = 0.79$), (5) Mean (across all surrogate representations) of the trace of the adjacency matrix of the network representation of the tuple - (HR signal, activity count signal, DC signal) at the 10th coarse-graining scale ($\mu_{TRACE(10)}$) ($\rho = 0.76$), (6) Variance (across all surrogate representations) of transfer entropy from the activity counts signal to the deceleration capacity signal at the 1st coarse-graining scale ($\sigma_{\tau_{Act \rightarrow DC}^{(1)}}^2$) ($\rho = 0.47$), (7) Variance (across all surrogate representations) of transfer entropy from the activity counts signal to the HR signal at the 1st coarse-graining scale ($\sigma_{\tau_{Act \rightarrow HR}^{(1)}}^2$) ($\rho = 0.31$), (8) Mean (across all surrogate representations) of transfer entropy from the activity counts signal to the HR signal at the 4th coarse-graining scale ($\mu_{\tau_{Act \rightarrow HR}^{(4)}}$) ($\rho = 0.29$), (9) Variance of the 5-minute PNN50 measure i.e., the average number of pairs of adjacent beat-

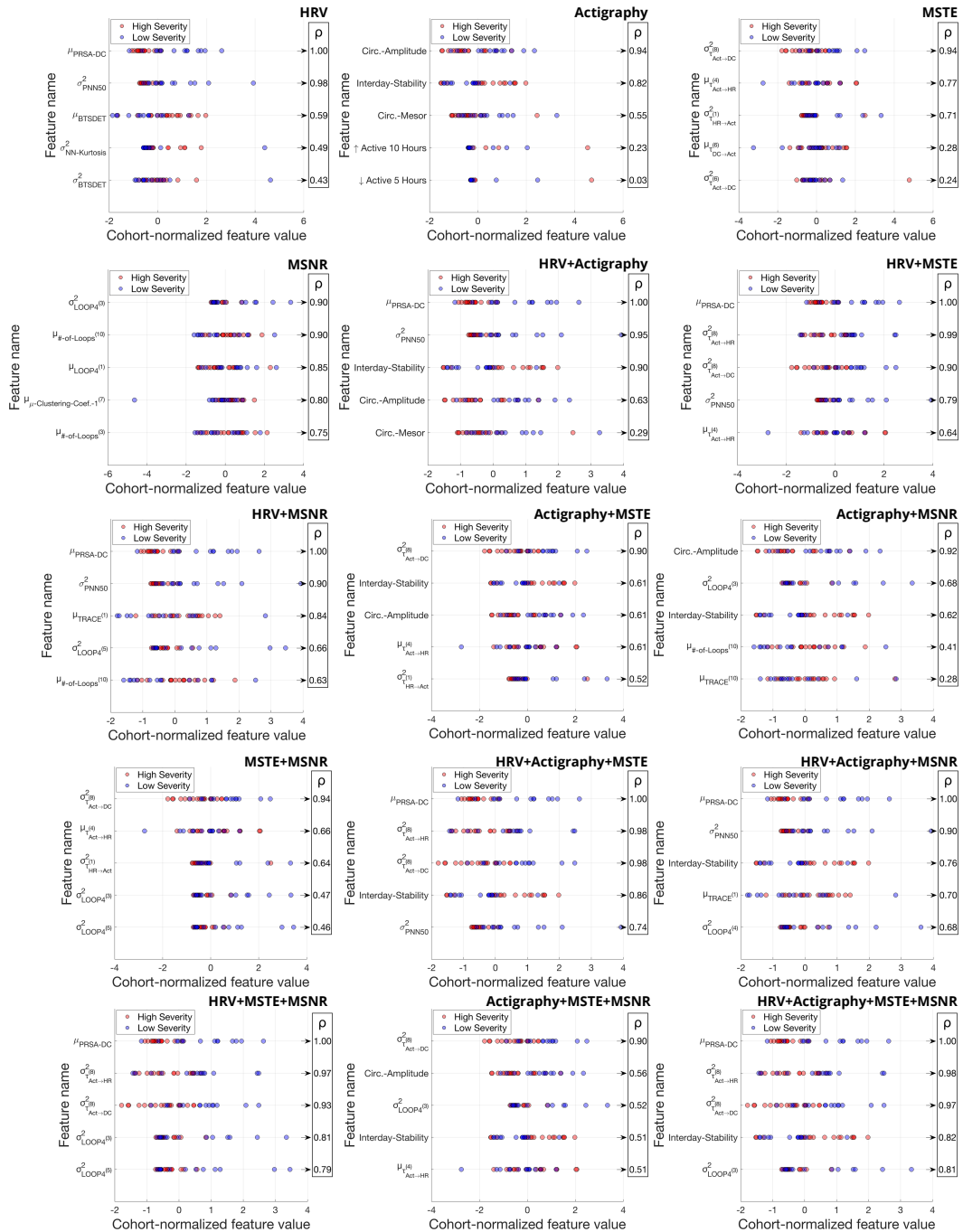


Figure 3.5 The top-five most popular features for each of the 15 classifiers.

to-beat intervals differing by more than 50 ms, between the times 10 PM and 10 AM (σ_{PNN50}^2) ($\rho = 0.18$). The feature weight for ($\mu_{PRSA-DC}$) was greater than zero for 93% of the time, suggesting an inverse relationship (due to the negative-log relationship between features and odds) between the feature values and the Rett disease severity (i.e., a higher values of $\mu_{PRSA-DC}$ corresponded to a higher chance of low-severity Rett syndrome). It conformed with our feature values as shown in Figure 3.4B. In Figure 3.5 we illustrate the top 5 most popular features for each of the 15 classifiers. Whenever HRV-metrics were used for classification (8 of the 15 instances), $\mu_{PRSA-DC}$ was the most important feature with a consistent feature popularity score of 1. When only Actigraphy-metrics were used for classification, the amplitude of the circadian rhythm measured using the cosinor rhythmometry was the most popular feature with a feature importance score of 0.94. When we used MSTE-features alone, $\sigma_{\tau_{Act \rightarrow DC}}^{(8)}$ was the most popular feature with a feature popularity score of 0.94. Finally, when we used the MSNR features alone, ($\sigma_{LOOP4(3)}^2$) was the most popular feature with a feature popularity score of 0.90. Interestingly, the top features from the individual feature set models for HRV-metrics ($\mu_{PRSA-DC}$), MSTE-features ($\sigma_{\tau_{Act \rightarrow DC}}^{(8)}$), and MSNR-features ($\sigma_{LOOP4(3)}^2$) were all featured as one among the top-5 most popular features in the best model (AUC = 0.92) that used the feature combination of HRV-metrics, MSTE-features, and MSNR-features. For the classifier that used all available features (HRV + Actigraphy + MSTE + MSNR), the following metrics featured as the top-5 most popular features: (1) $\mu_{PRSA-DC}$ ($\rho = 1.00$), (2) $\sigma_{\tau_{Act \rightarrow HR}}^{(8)}$ ($\rho = 0.98$), (3) $\sigma_{\tau_{Act \rightarrow DC}}^{(8)}$ ($\rho = 0.97$), (4) Interdaily Stability ($\rho = 0.82$); (5) $\sigma_{LOOP4(3)}^2$ ($\rho = 0.81$).

3.4 Discussion

As of 2022, no clinically meaningful disease-modifying treatments exist for patients with Rett syndrome. We instead rely on multiple therapeutics and symp-

omatic treatment strategies geared towards managing respiratory ailments, treating seizures, improving gastrointestinal function, and improving motor skills [129]. As new drugs and therapeutics are discovered, the need for objective measures that can be used in clinical trials increases. Neurological disorders, including Rett, suffer from difficult-to-measure symptoms. Most efficacy assessments are based on parent, clinician (and in some cases, patient) interpretation of symptom severity, which by nature introduces bias and often, exemplified by high placebo rates, results in a high noise to signal ratio. Hence, the need to establish objective measurements in patients directly, especially for symptoms that are difficult or impossible to observe, would open the door to evaluate therapeutics in novel ways and has the potential to expedite therapeutic development in multiple ways. Namely, (1) It would reduce bias; (2) It would help reduce clinical trial sample size by reducing the noise to signal ratio; and (3) It would facilitate shorter trial duration by capturing continuous data at home. The measurement of autonomic function could be an early biomarker of therapeutic efficacy and may be particularly relevant for curative therapeutics such as gene therapies that theoretically should improve or restore baseline function. If trials focus on efficacy measures that require learning and implementation (like mobility, communication, and hand use), this may take significantly longer to detect than an autonomic function correction that should not require learning. Thus, in the current study, we attempted to address these unmet needs by capturing physiological (ECG) and body activity (three-axis accelerometer data) from a 20-patient cohort. We chose to regress features extracted from the ECG signal and body activity measurements against the binarized CGI-S to correlate objective measures attained from a wearable biosensor to an overall symptom severity scale. We have shown that the inclusion of multiscale features (MSTE and MSNR) along with HRV-metrics provided a performance improvement of 21% in terms of the AUC ($AUC = 0.92$) when compared to using HRV-metrics

alone (AUC = 0.76). This improvement is attributable to the information embedded in the higher-order interactions between the HR, DC, and activity count signals. While the transfer entropy-based features enabled us to study the 2-dimensional variations between all combinations of HR, DC, and activity count signals in both directions, the network representations-based features enabled us to study the 3-dimensional interactions between these signals. Further, the computations of these features at multiple coarse-graining scales provided a means to analyze the signal modulations occurring due to different physiological phenomena (at different timescales). The coarse-graining scales we used in this study ranged from 1 to 10, which corresponded to a variation of sampling rate from 1/30 Hz to 1/300 Hz (i.e., from a sample every 30 seconds to a sample every 5 minutes). It allowed us to study the effect of different physiological processes, including respiration, vagal activity, and circadian rhythm, on the interactions between HR, DC, and activity count signals.

Our analysis suggested that the mean value of the DC of the HR signal captured on the BioStamp[®] nPoint between 10 PM and 10 AM was the most popular feature to classify low-severity Rett syndrome patients from high-severity Rett syndrome patients. It was the most popular feature in all the eight classifier models in which it was used, with a consistent and highest feature popularity score of 1. The computation of DC involved synchronizing the phases of all periodic components of the signal irrespective of their timescales [130]. Thus, the DC captured the overall deceleration of the sinus rhythm due to physiological processes that occurred at different timescales, including the vagal (parasympathetic) activity, the baroreflex, and the circadian rhythm. In Figure 3.4B, we depicted the variation of the 5-minute averages of DC between the times of 10 PM and 10 AM for both low-severity and high-severity Rett syndrome patient visits. Apart from the mean value of the DC, we observed the emergence of the variance of transfer entropy at 8th coarse-graining

scale from (1) Activity count signal to the HR signal and (2) Activity count signal to the DC signal among the most popular features. Further, the variance of the number of loops of size 4 in the network representations at coarse-graining scales 3 and 5 were among the top-5 most popular features in the best classifier model (AUC = 0.92). Thus, in Fig Figure 3.4C, we illustrated an example of the network representations for a low-severe and a high-severe Rett patient at the 3rd, 5th, and 8th coarse-graining scales. In addition to demonstrating the viability of a wearable biosensor to estimate disease severity based on objective measurements directly in patients, another novelty in our work was handling missing data in signals captured using wearables for patient state analysis. We proposed three different techniques for this purpose. The first method of combining data from multiple channels boosted the amount of available data by 13.4% when compared to using a single channel (medial chest) and improved the average SQI by 10%. In Fig Figure 3.4A, we illustrated this improvement in SQI for each of the 32 patient-visits due to the usage of the novel SQI based ECG imputation. The second method helped impute the activity count data and reduce missingness. The third and final technique of generating stochastic surrogates enabled us to compute the MSTE and MSNR features. It was specifically developed for this study, without which no multiscale analysis could have been performed. We computed the sample mean and sample variance of all MSTE and MSNR metrics across all surrogate representations and used them as features for our classification models. The sample mean is a measure of the samples' central tendency, and the sample variance is a measure of the spread of the samples. Since our method used a normal distribution to sample the random variables, and a normal distribution can be completely defined by a mean value and variance, it was rational to use the sample mean and variance to characterize the underlying MSTE and MSNR values.

In section 2.3, we listed the different medications prescribed to the Rett patients

in our study. Our literature survey yielded that these drugs, prescribed to minimize the symptoms of Rett syndrome have been shown to partly improve autonomic cardiovascular function. Any suppression in HRV metrics for the high severity group, if observed in our experiments, could imply that high severity Rett patients inherently are less adaptable to the autonomic changes due to the severity of the disease.

3.5 Conclusion

We developed a machine learning model that utilized features characterizing HRV, body movement, and the interaction between the two to estimate Rett syndrome severity in a group of 20 female Rett patients. For this, we developed a novel stochastic method for biosignal data imputation. We obtained the highest pooled-AUC equal to 0.92 utilizing the feature combination of HRV-metrics, MSTE-features, and MSNR-features. Further, the proposed approach provided us with physio-motor biomarkers that could be used in clinical trials as objective metrics to quantify the improvement in overall patient state. Specifically, the mean DC of the HR signal captured between 10 PM and 10 AM using the BioStamp[®] nPoint biosensor was the most popular feature with a feature popularity score equal to 1. In conclusion, our study (1) Implemented a novel data imputation technique for physiological signals, (2) Developed a machine learning model to estimate Rett disease severity, and (3) Developed objective measures that characterize the autonomic function in Rett syndrome.

CHAPTER 4

PATIENT STATE ANALYSIS FROM NEARABLES

4.1 Introduction

In this chapter, we describe an edge computing and ambient data capture system for clinical and home environments. Further, we will describe five applications of the data capture system. In clinical environments such as ORs and intensive care units (ICUs), key events during patient monitoring include: (1) Patient movements while lying in bed and in mobility within the room [131, 132]; (2) Bedside monitor alarm triggers and noise pollution [132, 133, 134, 135, 136]; (3) Presence, absence and movement of clinical personnel in the patient's vicinity [137, 138, 139]; and (4) Variations in the ambient light, temperature, and humidity [132, 140, 136]. In home environments, key events that are generally untracked but are beneficial for patient monitoring include: (1) Patient bodily movement during sleep [141, 142]; (2) Patient movement around their residence [143]; (3) Doorbell triggers, smoke-detector triggers, microwave beeps, and phone rings [144]; and (4) Changes in the ambient light, temperature, and humidity [145]. Recently, nearables [16, 87, 145, 146, 147, 148] such as microphones, video cameras, light-intensity sensors, temperature and humidity sensors, are becoming more popular for hassle-free patient monitoring. They not only collect valuable patient behavior data but also pick up key information about the patient's ambient environment while not interfering with the patient's day-to-day activities. Thus, we developed a non-contact data capture and archival system to capture patient behavior and ambient environment information.

While obtaining patient behavior and ambient information is crucial in understanding the effects of healthcare practices on patient health, maintaining patient

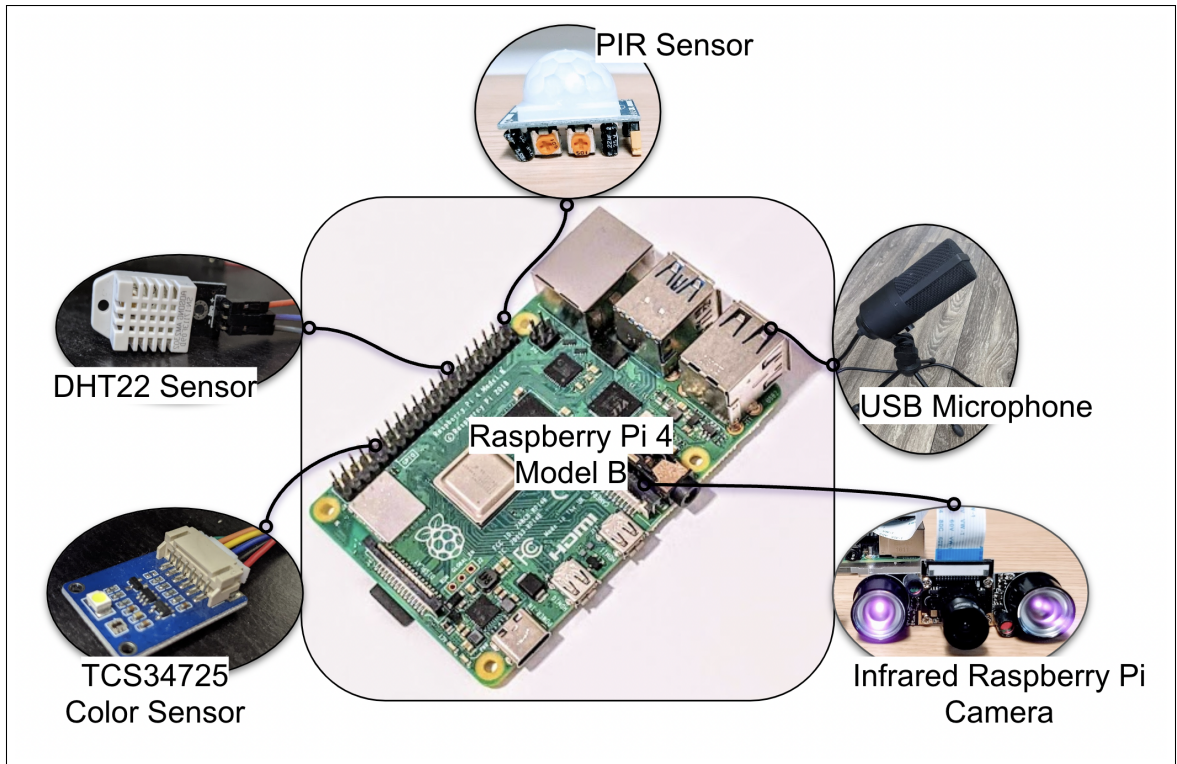


Figure 4.1 Edge computing and ambient data capture system. (PIR: Passive Infrared Sensor; USB: Universal Serial Bus)

privacy is as important if not more. For this, we utilized the edge computing paradigm. In *edge computing*, algorithms are decentralized and moved closer to the point of data capture to reduce latency and bandwidth requirements. This paradigm can be defined as computing outside the cloud, happening at the edge of the network, specifically in applications where real-time data processing is required. In our work, we utilized a Raspberry Pi (RPi) as a hub for data collection and edge computing. We extracted patient privacy-preserving features from the captured data on the RPi (at the edge) before discarding the raw underlying signal and transferred the computed features to a Health Insurance Portability and Accountability Act (HIPAA) compliant storage. The retrospective patient state analysis utilizing these captured features occurred on a central server away from the RPi. Non-contact monitoring of patients is becoming more prevalent, especially

in elderly patients [149, 150, 151] and neurodevelopmental populations (such as Autism Spectrum Disorder) [152] as these systems cause no burden on the patients in terms of wearing and operating the device (in contrast to a wearable such as the smart watch). Further, non-contact monitoring allows for monitoring the patient's global movements in contrast to wearables, thus providing additional information about patient behaviors. The advent of the COVID-19 pandemic further increased the need for such systems [147] as they allow passive patient monitoring from a distance. However, there are multiple challenges in building such a system. First challenge lies in integrating different sensors to capture multiple data modalities under a single clock. Second, this system should asynchronously transfer the data to a HIPAA-compliant database. Lastly, the system should maintain patient privacy while capturing the various data modalities. We overcame these challenges by developing a novel software system that ran on an RPi. Using this system we integrated the following five sensors: (1) PIR sensor (2) RPi-Infrared (IR) camera (3) Universal Serial Bus (USB) Microphone (4) TCS34725 color sensor and (5) DHT22 temperature-humidity sensor. Further, we utilized the onboard Bluetooth receiver to geolocate humans using Bluetooth beacons. The main novelty of our work lied in *capturing privacy-preserved features simultaneously from multiple sensors to perform human movement detection, registering auditory cues, human geolocation, and ambient environment monitoring*. It was a specific design consideration that all hardware could be easily purchased at a low cost. This effectively helped scale the system and enabled us to capture data in large clinical environments. Further, the system was intended to be deployed rapidly without the need for expert fabrication of hardware.

We have described four applications of the data captured by our system. Namely: (1) Medical equipment alarm classification, (2) Geolocation of humans in a built environment, (3) Ambient light logging, and (4) Ambient temperature and humidity

logging, (5) Obstructive sleep apnea detection from PIR sensor, (6) Next Day Behavior Prediction in Autism. We provide the details of these experiments in section 4.3. While alarm classification and geolocation tasks aid us in non-contact patient monitoring, the management of indoor air temperature, humidity, and light intensity are vital for maintaining patients' comfort in hospitals [153]. In [154], researchers recommended creating separate thermal zones according to patients' preferences for improved recovery. Further, a relative humidity (RH) greater than 45% – 50% assists fungal growth in built environments and can affect patient recovery in hospitals [155]. With respect to the lighting conditions, patients prefer a mixture of warmer and cooler luminaire throughout the day and favor distributed lighting over the traditional over-bed configuration [69]. Thus, temperature, humidity, and light intensity monitoring are essential components in our ambient data capture system. These sub-parts assist clinicians and caretakers make critical decisions about the patient's environmental conditions.

4.2 Materials and methods

4.2.1 System architecture

The system architecture described here is a low-cost, high-compliance design. At its core is the Python script that interfaces with the sensors via the RPi. A picture illustrating the hardware components of the system is shown in Figure 4.1. The bill-of-materials, along with the total cost for the hardware components and the system dimensions, are provided in our open-sourced `GitHub` repository [156]. We now describe the system's individual hardware components along with the associated software.

Raspberry Pi

The RPi is a \$35 computer that is about the size of a deck of cards. It functioned as the central hub in our data collection pipeline. In our work, we used the RPi 4 model B (Figure 4.1), which was released in June 2019 and was the most recent model during the software development stage of the project. The Debian-based operating system ('Raspbian-Buster') that is optimized for the RPi was installed for developers and users to interact with the hardware. Among the onboard peripherals on the RPi were two USB 2.0 ports and two USB 3.0 ports, a 40-pin General Purpose Input-Output (GPIO) header, and a USB-C port to supply power to the RPi. The RPi was powered using a 5V 3A power adapter. Alternatively, in low-resource settings, one can consider powering the RPi using alternative means, for instance, a car battery [147]. In that case, a step-down transformer is required to convert a 12V power supply to 5V. However, in all our experiments, we used an external power supply that was available in the built environment.

PIR sensor based human movement detection

We used a PIR sensor (Figure Figure 4.1) for coarse human movement detection. The PIR sensor consists of a pair of IR sensitive slots housed in a hermetically sealed metal casing. A Fresnel lens acts as the outermost cover, which increases the range and sensitivity of the sensor. When the sensor is idle, both the IR slots receive the same amount of IR radiation. Whereas, when an IR emanating object moves past the field of view of the first IR slot, this slot detects an increased IR radiation, and thus, a differential signal C between the two slots is generated. A differential signal C' , which is completely out of phase with respect to C , is generated when the object moves past the other IR-sensitive slot. These differentials are then processed to form the output signal. Our system was designed to capture data at a sampling frequency of 1Hz. The data itself was a binary spike train taking the value 1 when a movement

was detected and 0 otherwise. In chapter 6, we will present a method for capturing movement data using a PIR sensor during sleep from a single person. Further, we show that the captured movement data can be utilized to build a binary classifier for obstructive sleep apnea classification at a classification accuracy equal to 91% [142]. One should note that this method can be generalized to perform human movement detection in different clinical and home environments. Based on the positioning and orientation of the sensor, one could capture different information. For instance, the timestamps when the patient’s knee was operated on could be obtained by placing the sensor to monitor a patient’s knee during their knee replacement surgery.

IR camera-based human movement detection

It is possible to use a video feed [16] from an RPi-IR camera in place of the PIR sensor to perform human movement detection. This method allowed us to capture the human movement signal with more than two quantization levels and obtain a finer signal than the binary signal captured using the PIR sensor. Besides capturing the occurrence of movements, the RPi-IR camera-based analysis allowed us to compute the intensity and direction of these movements. We used the No Infrared (NoIR) version [157] of the RPi-IR camera (Figure 4.1). In contrast to the regular RPi-camera, the NoIR RPi-IR camera did not employ infrared filters and gave us the ability to see in the dark with infrared lighting.

We now describe the method to extract human movement signal from the RPi-IR camera feed and describe its utility. Without loss of generality let us assume we capture the video at 1 Hz (i.e. 1 frame per second). Let the video frame at time t (in seconds) with a pixel resolution equal to $M \times N$ be denoted as F_t , and the previous video frame be F_{t-1} . The frame difference between the two frames at time t (D_t) is defined as $D_t = F_t - F_{t-1}$. The difference-frame D_t has the same pixel resolution as F_t and F_{t-1} , i.e. $M \times N$. For a given video V , the corresponding difference frame-

stack is given by the set $D = \{D_t\}_{t \in [2, T]}$ where T is the total number of frames in V and D_t is the difference frame at t seconds. We extract four different signals from D namely: (1) Global Difference Sum (*GDS*); (2) Global δ -Pixel Count (*GDPC*); (3) Local Difference Sum (*LDS*); and (4) Local δ -Pixel Count (*LDPC*). The global signals (*GDS* and *GDPC*) for a given difference frame D_t are computed as follows:

$$GDS[t] = \sum_{i \in [1, M]} \sum_{j \in [1, N]} |D_t[i, j]| \quad (4.1)$$

$$GDPC[t] = \#(\{d \in D_t \mid |d| > \delta\}) \quad (4.2)$$

where d denotes an individual pixel in the difference frame D_t , $|\cdot|$ denotes the absolute value and $\#(\cdot)$ denotes the set cardinality. $GDS[t]$ is the sum total of the absolute values of the pixels in the difference frame at time t seconds and $GDPC[t]$ is the total number of pixels in the difference frame at time t seconds that have an absolute value greater than δ . If $\delta = 0$, $GDPC[t]$ denotes the total number of non-zero pixels in the difference frame D_t . The local signals (*LDS* and *LDPC*) for a given difference frame D_t are computed likewise to the global signals but are calculated on smaller blocks in the difference frame. For this, we divide a difference frame of pixel resolution $M \times N$ into K parts of equal size along the M -axis and L parts of equal size along the N -axis. Each of the K parts contains $m = \frac{M}{K}$ columns and each of the L parts contains $n = \frac{N}{L}$ rows. This division along the rows and columns of a difference frame D_t creates $nBlocks = K * L$ local-blocks of size $m \times n$. Let the s^{th} local-block be denoted as $DL_{t,s}$. Then the local signals (*LDS* and *LDPC*) at time t for the s^{th} local-block are given as:

$$LDS[t, s] = \sum_{i \in [1, m]} \sum_{j \in [1, n]} |DL_{t,s}[i, j]| \quad (4.3)$$

$$LDPC[t, s] = \#(\{d \in DL_{t,s} \mid |d| > \delta\}) \quad (4.4)$$

where d denotes an individual pixel in the difference frame local-block $DL_{t,s}$, $|\cdot|$ denotes the absolute value and $\#(\cdot)$ denotes the set cardinality. The $LDS[t,s]$ is the sum total of the absolute values of the pixels in the s^{th} local-block of the difference frame at time t seconds and the $LDPC[t,s]$ is the total number of pixels in the s^{th} local-block of the difference frame at time t seconds that have an absolute value greater than δ . Similar to $GDPC$, if $\delta = 0$, $LDPC[t,s]$ denotes the total number of non-zero pixels in the s^{th} local-block of the difference frame D_t . The different parameters that needed to be set were δ, K , and L . The default values we set in our work were $\delta = 0, K = 5, L = 4$ and the videos we experimented on had a pixel resolution equal to 320×256 . Thus, in our work, we had $m = 64, n = 64$, and $nBlocks = 20$.

Together, the host of time-series data described above gave us information about the temporal and spatial variations in the video V . We needed certain assumptions so that these signals gave us information about human movement. The assumptions were as follows (1) We had one person in the entire video. (2) The background was static, and the only moving object in the video was the person or an object attached to the person. (3) The video was not corrupted or affected by noise.

While the PIR sensor provided binary movement signals (movement occurred vs. no movement), the movement signals from the IR cameras were finer and had a higher spatial resolution. These two solutions were useful in different scenarios. For instance, while the PIR sensor could be used to detect the presence or absence of a human in a room, the movement signal from the IR camera could be used to perform a privacy-preserving analysis of a patient's sleep. Thus, based on the requirement and budget, the users of our system could use one or both sensors. For instance, if the application is human movement detection and cost is a concern, we recommend the users to use a PIR sensor.

Privacy preserving audio data capture

The proposed audio data capturing system consisted of a USB microphone connected to the RPi. In this work, we tested the *Fifine Conference USB Microphone* (Figure 4.1). In order to record the audio signal, we used the *python-sounddevice* package available on `GitHub` [158, 159] under the MIT License. Specifically, we modified the script *rec_unlimited.py* [158] to be able to continuously record audio data and perform audio feature computation at regular intervals. To extract features from these audio snippets, we used a 30 millisecond Hanning window with a 50% overlap. Feature computation was done using the *librosa* package available on `GitHub`[160] under the Internet Systems Consortium (ISC) license. We utilized the spectral representation method *stft* in the Core IO and Digital Signal Processing toolbox and the spectral features method *mfcc* in the feature extraction toolbox in order to compute short-time Fourier transform (STFT) and Mel-frequency cepstral coefficients (MFCC), respectively. Further, we used the filter bank construction method *mel* in the Filters toolbox to create a filter bank with 10 frequency bins. For each 30 millisecond window, we then computed the signal energy in different frequency bins by performing the following matrix multiplication:

$$E = MS \tag{4.5}$$

where S was the STFT coefficient vector for the current window, M was the mel filter mask matrix with each row corresponding to a different mask, and E was the signal energy in different frequency bins corresponding to the mel filter masks. Note that the default shape of mel filter masks was a triangle with the mask values summing to one. We further included scripts for computing sample entropy of the windowed audio snippets using the *mse.c* script, which was available on `Physionet` [161] under the GNU general public license. Further, we developed scripts to

archive the computed audio features to a secure cloud storage and discard the underlying audio snippets. The above implementations were developed using Python 3.7.3, C, and Bash scripting. These were representative edge computing methods that could extract different features (MFCC, signal energy and sample entropy) from audio signals. Other feature extraction algorithms which could be run on the constrained environment of an RPi could be easily incorporated as additional methods. The discarding of raw audio data ensured patient privacy and speaker identification was not possible. Furthermore, we did not record the speaker's pitch information in the audio snippets or deploy methods to determine if a given window contained voiced speech.

Human location tracking via Bluetooth

The Bluetooth scanning system utilized the onboard Bluetooth receiver of an RPi. In this work, we tested the *Smart Beacon SB18-3* by *contact.io* with an RPi 4 model B. We leveraged the *scanner* package by *bluepy - a Bluetooth LE interface for Python* [162] for this purpose. The code implementation was done in Python 3.7.3. The software recorded the received signal strength indicator (RSSI) value from all the beacons transmitting the Bluetooth signal in the vicinity. We used the media access control (MAC) addresses of the Bluetooth beacons to identify them. A Python script would poll for RSSI values from all the beacons in the vicinity at regular intervals. The received RSSI values, the unique MAC address, and the recording timestamp were dumped into a file.

Ambient light intensity assessment

In this work, we used the *Waveshare TCS34725 Color Sensor* (Figure Figure 4.1) in conjunction with an RPi to capture ambient light intensity. Among other signals, the color sensor captured the following signals which were of interest to us: (1)

Red, Green, and Blue values in the RGB888 format (8-bit representation for each of the three color-channels); (2) Illuminosity in Lumen per square meter (LUX); and (3) Color Temperature in Kelvin. The RGB values and the color temperature gave us information about the ambient light color, and the illuminosity values gave us information about the ambient light intensity.

Temperature and humidity detection

For temperature and humidity detection we used the DHT22 Temperature-Humidity sensor module (Figure Figure 4.1) in conjunction with an RPi. The DHT22 sensor comprised a thermistor and a capacitive humidity sensor that measured the surrounding air to provide calibrated temperature and humidity values. Further, the sensor module came with a digital board that housed three pins, namely VCC, GND, and OUT. The sensor had an operating voltage of 3.3/5V (DC), and the OUT could be read from a GPIO pin on the RPi. The temperature range was -40 to 80 °C, and the humidity range was $0 - 100\%$ RH.

4.2.2 Data fusion

The following data modalities were captured at a sampling frequency of 1 Hz: (1) PIR sensor-based human movement; (2) IR camera-based human movement; (3) Audio data; (4) Bluetooth RSSI signal; (5) Ambient light intensity; and (6) Temperature and humidity. A single main script facilitated the capture of all the above data modalities and the individual time stamps corresponding to each sample. The human pose signal was recorded in an ad-hoc manner when the algorithm detected a human body. Nevertheless, the corresponding timestamps were recorded using a single clock onboard the RPi to ensure all data modalities were recorded synchronously. Further, the recorded data was easy to access via a simple directory structure consisting of separate folders for each data modality. The

data collection was performed in parallel by each RPi and transmitted in real-time to a HIPAA compliant central server, which aggregated the data to perform patient state analysis.

4.3 Applications

The applications of our edge computing and ambient data capture system range from monitoring patient sleep in sleep labs to tracking neurodevelopmental disorder patients at their homes. In our work, we describe four experiments to demonstrate our system's utility:

1. Obstructive sleep apnea classification using passive infrared sensor: *We can detect the binarized apnea hypopnea index based on the body movements during sleep via the PIR sensor.*
2. Medical equipment alarm classification using audio: *This utility facilitates patient monitor alarm monitoring in ORs or ICU rooms, where there are many systems not centrally integrated.*
3. Geolocation of humans in a built environment: *We can track humans in a built environment and model social distancing for quantifying epidemic disease exposure [163].*
4. Ambient light logging: *This system can be used to study the effect of ambient light on human circadian rhythm.*
5. Ambient temperature and humidity logging: *We can perform long-term monitoring of the effects of ambient environmental conditions on patient behavior.*
6. Next Day Behavior Prediction in Autism: *We can predict daily behaviors of people with Autism based on their previous night's sleep.*

4.3.1 Obstructive sleep apnea classification using a passive infrared sensor

In this experiment, we illustrated the use of the PIR sensor (a nearable) for patient state analysis. Specifically, we used the data obtained from the PIR sensor to detect patients with obstructive sleep apnea.

Dataset

We used a subgroup of participants included in the Emory Twin Study Follow-up and recruited from the Vietnam Era Twin Registry [164]. Overnight sleep movements of participants were continuously recorded using the low-cost PIR sensor in our nearables system, for up to eight hours during a conventional laboratory-based PSG. Breathing was measured with separate channels for oral/nasal airflow, nasal pressure, thoracic and abdominal respiratory effort, and pulse oximetry. All data were collected at Emory sleep-labs and data acquisition systems were setup in two bedrooms present in the sleep-labs. We obtained written informed consent from all participants, and the Emory University institutional review board approved this research (IRB#00081004). Data were collected between January 2016 and October 2018 during simultaneous polysomnogram studies on 32 male participants (aged $\in [61$ years, 73 years]). 14 of the analyzed participants were classified as suffering from OSA while 18 were deemed non-OSA (control) participants. The Apnea-Hypopnea Index (AHI), which is an indicator of the severity of OSA measured during the PSG, was used to determine OSA. Participants with an $AHI \geq 15$ events per hour were assigned to the OSA class. Those with $AHI < 15$ served as the control group. The sample used in this study included 11 participants demonstrating PLMD, defined as having a Periodic Limb Movement Index (PLMI) – the number of periodic limb movements per hour as recorded by expert annotators - greater than 15 movements per hour. The range of values for PLMI in the sample used in this study was $[0, 80.3]$.

Feature extraction

Overnight sleep movement data corresponding to i^{th} participant P_i during a single night recording was a summation of L_i time delayed impulse signals. This record was denoted as R_i . The sampling frequency was set to 1 Hz. Thus, for the i^{th} record R_i acquired between 10 pm in the night and 6 am the next day (eight hour recording), we had a maximum of $8 \times 3600 = 28800$ impulses i.e. $\max_{i \in \mathcal{N}} L_i \leq 28800$, where \mathcal{N} was the number of participants in the data cohort. We concatenated all the timestamps in Portable Operating System Interface time (POSIX-time) corresponding to impulses present in record R_i to construct i^{th} POSIX-time timeseries T_i . We then defined the i^{th} first-difference signal (D_i) as the first difference of T_i and the i^{th} activity signal (A_i) as the sample-wise inverse of this first difference. Since the sampling frequency was set to 1 Hz, the possible minimum time span between two movement events was one second, whereas the maximum time span was finitely large. This provided a natural way for normalizing the activity signal and it was always bounded between 0 and 1, i.e. $A_i[t] \in [0, 1] \forall t \in [10 \text{ pm}, 6 \text{ am}]$.

We utilized a total of 12 features to perform OSA classification. The 12 features belonged to one of the following three categories 1) Entropy based features; 2) Standard statistics of the first difference of the the time series of movements; and 3) Periodic limb movement activity. Multiscale entropy (MSE) is a method to measure complexity of a physiological or a finite time series with the aim to consider multiple time scales inherent in such time series [165]. In order to utilize sleep movement signal complexity in OSA classification the first set of features were based on MSE, denoted \mathcal{H}_τ where τ is the scale ($\tau \in \{1, 2, 3, 4, 5\}$). The Difference Signal Statistics (DSS) were derived from the first difference signal D_i . The number of impulses in the recording, i.e. number of samples in the difference signal plus 1, was a feature. The mean, variance, skewness, kurtosis and inter-quartile range of the difference

signal were the other features in this second set. PLMI measured as part of the sleep study polysomnography for each of the 32 participants was an additional feature.

Experiments

We performed six experiments to classify each participant to one of the OSA or control class where in each experiment we used a different combination of features. Specifically, the six feature combinations were as follows: (1) MSE; (2) DSS; (3) PLMI; (4) MSE + DSS; (5) MSE + PLMI; and (6) MSE + DSS + PLMI. For each of these feature combinations we further implemented the forward selection method [166] to simultaneously rank and select the most discriminative features for the classification task. Logistic regression was used as the classifier and Leave One Out Cross-Validation (LOOCV) was used to assess the performance. In each iteration of the cross-validation, data corresponding to one participant was left out as the testing data point and \mathcal{P} classifiers were built using the remaining data (utilizing the remaining 31 participants' data as the training data) where \mathcal{P} corresponds to the total number of features in the analysis and is defined as follows:

$$\mathcal{P} = \begin{cases} 1 & \text{for } 3^{rd} \text{ experiment} \\ 5 & \text{for } 1^{st} \text{ experiment} \\ 6 & \text{for } 2^{nd} \text{ and } 5^{th} \text{ experiments} \\ 11 & \text{for } 4^{th} \text{ experiment} \\ 12 & \text{for } 6^{th} \text{ experiment} \end{cases}$$

The p^{th} classifier ($p \in [1, \mathcal{P}]$) corresponded to the model trained with best validation accuracy when using only ' p ' of the \mathcal{P} features. The performance metrics included the F1-Score (\mathcal{F}_1) and accuracy (Acc) and were calculated and reported for the model with the highest \mathcal{F}_1 .

Table 4.1: Class labels including the Musical Notes in International Organization for Standardization / International Electrotechnical Commission (ISO/IEC) 60601 – 1 – 8 Alarm and their count [*E* stands for the *Empty* class and *T* stands for the *Transition* class]

Musical Note	Fundamental Frequency (Hz)	Count
C_4	261.63	207
D_4	293.66	43
E_4	329.63	32
F_4	349.23	46
$F_4^\#$	369.99	27
G_4	392.00	66
A_4	440.00	45
B_4	493.88	17
C_5	523.25	81
<i>E</i>	-	969
<i>T</i>	-	432

4.3.2 Medical equipment alarm classification using audio

In this experiment, we tested the utility of the audio feature extraction methods (energy in mel frequency bins and MFCC) proposed by us for clinical audio classification. For this, we utilized an external clinical audio database, extracted the proposed features, and performed multi-class classification.

Dataset

We analyzed the International Organization for Standardization / International Electrotechnical Commission (ISO/IEC) 60601 – 1 – 8 type medical equipment alarm sounds [167]. The alarm sounds comprised eight categories: general, oxygen, ventilation, cardiovascular, temperature, drug delivery, artificial perfusion, and power failure. Further, each category had two alarm sounds, namely, medium priority alarm and high priority alarm. The medium priority alarm sounds were about one second long and consisted of three musical notes that were played once, whereas the high priority alarm sounds were about 4.5 seconds long. They consisted

of five musical notes that were played twice. All the alarm audio recordings were single channel, sampled at 22050 Hz, and recorded in the *Waveform* audio file format. Nine musical notes were used to construct these 16 different alarm audio recordings. Table Table 4.1 lists these nine musical notes with their fundamental frequencies. The works [167, 168] provide more information on the individual alarm sound recordings.

Feature extraction

We used the audio data capture software described in Section subsection 4.2.1 and computed 20 MFCC features and 10 filter bank energy features on 30 millisecond snippets of the 16 alarm sound recordings. Further, we computed STFT coefficients for the audio clips and annotated each snippet to belong to one of the following 11 classes: $\{Empty, C_4, D_4, E_4, F_4, F_4^\#, G_4, A_4, B_4, C_5, Transition\}$. The *Empty* class was assigned when all the STFT coefficients of a snippet were equal to zero. If a particular audio snippet was partially made up of a specific note with the rest of the samples equal to zeros, such windows were annotated as the *Transition* class. We annotated the musical notes by comparing the fundamental frequency in STFT with the values shown in Table Table 4.1. Moreover, we used the note sheets provided in [167, 168] to confirm our annotations. We had a total of 1965 data points. Table Table 4.1 further provides the breakdown of the number of data points in each class.

Classification

Using the 30 features described, we performed an 11 class classification using five-fold cross-validation and an XGBoost [169] classifier. All codes were written in Python 3.6.3 and XGBoost was implemented using the package provided in [170]. The following hyperparameters were used without any tuning: $n_estimators = 150$,

objective = 'multi:softmax', num_class = 11, max_depth = 6. All other hyperparameters were set to their default values. As illustrated in Table Table 4.1, the dataset contained class imbalance. We thus used both the macro averaged $\mathcal{F}1$ ($\mathcal{F}1 - macro$) score and the micro averaged $\mathcal{F}1$ ($\mathcal{F}1 - micro$) score as the measures for assessing classification performance. The $\mathcal{F}1 - macro$ score gives equal importance to each class irrespective of the number of samples in each class thus providing a balanced assessment of the multi-class classification performance when the dataset is imbalanced. The $\mathcal{F}1 - micro$ score on the other hand aggregates samples from all classes before computing the $F1$ score. Please refer [171] for the individual expressions for computing the two $F1$ scores.

Speech mixing

Next, we measured the performance of the note classification algorithm in the presence of speech. For this, in addition to the ISO/IEC 60601 – 1 – 8 dataset, we used a speech record consisting of five speakers [four male and one female] from the Oxford Lip Reading Sentences 2 dataset [172]. First, we resampled the speech record to match the sampling frequency of the alarm audio recordings (22050Hz) and extracted the first channel of this resampled speech record, denoted by \mathcal{S} . Next, for each of the 16 alarm audio records $\mathcal{A}_i, i \in [1, 16]$, we uniformly randomly pick an audio snippet from the speech record \mathcal{S}_i which was of the same length as \mathcal{A}_i . We then generated 10 audio records per alarm audio recording as follows:

$$\mathcal{M}_i[\alpha] = \alpha * \mathcal{S}_i + (1 - \alpha) * \mathcal{A}_i \quad (4.6)$$

where $\mathcal{M}_i[\alpha]$ was the mixed audio recording for a given α and $\alpha \in \{0, 0.1, 0.2, \dots, 0.8, 0.9\}$ was the mixing parameter that combined speech recording with alarm audio recordings. Note that $*$ and $+$ denoted scalar multiplication and sample-wise

Table 4.2: The human tracking experiment

Start Time (second)	End Time (second)	Duration (seconds)	Action
0	146	146	Stay in room 1
146	159	13	Move from room 1 to room 2
159	268	109	Stay in room 2
268	286	18	Move from room 2 to room 3
286	470	184	Stay in room 3
470	480	10	Move from room 3 to room 1
480	600	120	Stay in room 1

addition, respectively. When α was equal to 0, we had no speech component, and thus, $\mathcal{M}_i[0]$ was equal to the original alarm audio recording \mathcal{A}_i . As α increased from 0 to 0.9, the speech component in $\mathcal{M}_i[\alpha]$ increased linearly, and the alarm audio component decreases linearly.

We obtained a total of 160 different audio recordings (10 mixed audio recordings per clean alarm audio recording). We re-computed 20 MFCC features and 10 filter bank energy features for these 160 audio recordings using 30 millisecond Hanning windows and a 50% overlap and obtained a total of 19650 feature vectors. The ground truth labels for the feature vector at different α values were the same as those for $\alpha = 0$. Utilizing these 19650 feature vectors and corresponding labels, we performed an 11 class classification of musical notes using five-fold cross-validation and XGBoost [169] classifier. The hyperparameters were the same as it was when there was no speech mixing.

4.3.3 Geolocation of humans in a built environment via Bluetooth

In this experiment, we processed the RSSI signal received by the RPi to perform room-level location detection of humans using a Bluetooth beacon. We set up nine RPis in a built environment where each RPi was loaded with the software to capture RSSI values as received from a specific Bluetooth beacon. The built environment

consisted of three rooms, and three RPIs were present in each room, approximately equidistant from the center of the room. A *Kontakt.io* Bluetooth beacon with a unique MAC address was used in the experiment. A human subject carried the Bluetooth beacon in their clothes and moved around the space as illustrated in Table Table 4.2. Please note that this is a minor limitation of our system and the usage of a Bluetooth beacon (albeit inside one’s clothes) resulted in our system not being completely non-contact.

The processing of the collected RSSI values to perform geolocation of humans was done on a central server. This was because we collected data from multiple RPIs to perform geolocation. Once the data was transferred to the cloud from each RPI, we downloaded the data onto a central server and performed geolocation. The RSSI signals captured by each RPI were non-uniformly sampled. Hence, these signals were converted to a uniformly sampled signal with a sampling frequency equal to 1 Hz by filling missing data using the following equation.

$$\begin{aligned}
 RSSI_n[currentTime] &= RSSI_n[previous] \\
 &\times \max([\beta \times (currentTime - previous), 1])
 \end{aligned}
 \tag{4.7}$$

where, $n \in [1, 9]$ was the index variable to recognize RPIs, $RSSI_n$ was the vector of RSSI values captured by n^{th} RPI, $currentTime$ was the time (in seconds) at which we did not have a reading of the RSSI value, $previous$ was the closest predecessor time point (in seconds) to $currentTime$ at which we had a reading of the RSSI value, β was the decay parameter that controlled the rate at which the RSSI value decayed when RSSI values were missing and \max was a function to compute the maximum value in the input vector. The time difference ($currentTime - previous$) was expressed in seconds.

Further, any RSSI value less than -200 dBm was clamped to -200 dBm to have all RSSI values in a fixed range. We set β to 0.2, which corresponded to maintaining

Table 4.3: The ambient light tracking experiment

Day	Time (HH:MM)	Action
Clear Day (D_{clear})	07:32	Sunrise
	18:10	Sunset
Night-1 (N_1)	18:55	Lights-ON
	23:45	Lights-OFF
	01:00	Data upload start
	02:00	Restart data collection
Cloudy Day (D_{cloudy})	07:31	Sunrise
	16:58	Lights-ON
	18:11	Sunset
Night-2 (N_2)	23:42	Lights-OFF

the previous value for 5 seconds before the RSSI values were decayed when the RSSI values were missing. Further, we computed the average RSSI signal for each room by computing the mean value of the RSSI signals captured by the three RPi in each room. We used the softmax function to obtain a probability vector that gave the probability of the human subject with the Bluetooth beacon to be present in each of the three rooms at any given point in time. The averaging of RSSI values from multiple RPi receivers and the further usage of the softmax function significantly suppressed the effect of noisy RSSI samples.

4.3.4 Ambient light logging

To perform ambient data logging, we set up an RPi with the Waveshare TCS34725 color sensor in a built environment. The RPi was loaded with the associated software needed to capture the ambient light intensity values. Table Table 4.3 provides a timeline of events that occurred during ambient light data capture. The duration between 1 AM and 2 AM was reserved for data upload, and no data capture was performed during this period. The color sensor was set up in a place that received natural sunlight during the day and received light from light sources in the room

during the night. The lights in the room were turned ON when the natural sunlight was not adequate for a normal human lifestyle. The lights in the room remained ON until “sleep time” of the residents in the built environment when the lights were turned OFF. We recorded the ambient light data for two consecutive days. In parallel, we tracked the weather conditions of the data collection site and recorded the minute-to-minute local cloud cover information. With this setup, we studied the effect of cloud cover, sunrise and sunset times, artificial lights in the room, and buildings around the data collection site on the ambient light data captured by the color sensor. We divided the entire time period into 4 sections: D_{clear} , N_1 , D_{cloudy} , N_2 . Here, D_{clear} represents the day-period (sunrise to sunset) on the first day when the skies were clear, N_1 represents the first night (sunset to sunrise), D_{cloudy} represents the day-period (sunrise to sunset) on the second day when the skies were extremely cloudy (average cloud cover $\geq 80\%$), and N_2 represents the second night (sunset to sunrise). Since the cloud cover information during the night had little or no effect on the ambient light intensity in the room, in our analysis, we only used the cloud cover data tracked during D_{clear} and D_{cloudy} .

4.3.5 Ambient temperature and humidity logging

We validated the DHT22 temperature and humidity sensor against a commercially available sensor in this experiment. We set up an RPi with the DHT22 sensor in a built environment. The RPi was loaded with the necessary software to continuously capture temperature and humidity values and the associated coordinated universal time (UTC) timestamps. We collected the temperature (T_{RPi}) and humidity (H_{RPi}) values with this setup at a sampling frequency of 1Hz for three consecutive days (about 72 hours). Further, we set up the *ORIA mini Bluetooth Temperature-Humidity sensor* (a commercial sensor) beside our RPi setup and simultaneously performed temperature (T_{cs}) and humidity (H_{cs}) measurements using the

Table 4.4: Classification performance when different feature combinations were used

[Acc = Accuracy, \mathcal{F}_1 = F1 Score]. p^* : number of features needed in a given feature combination (with \mathcal{P}) to maximize \mathcal{F}_1 .

Experiments	Highest \mathcal{F}_1	Highest Acc (%)	p^*/\mathcal{P}	# of times each feature was picked for the highest \mathcal{F}_1 model of the 32 LOOCV-iterations												
				\mathcal{H}_1	\mathcal{H}_2	\mathcal{H}_3	\mathcal{H}_4	\mathcal{H}_5	L	μ_d	σ_d^2	γ_d	\mathcal{K}_d	I_d	$PLMI$	
MSE	0.81	84	3/5	31	13	18	2	32	-	-	-	-	-	-	-	-
DSS	0.62	66	4/6	-	-	-	-	-	5	21	30	32	28	12	-	-
PLMI	0.29	53	1/1	-	-	-	-	-	-	-	-	-	-	-	-	32
MSE + DSS	0.83	84	10/11	30	30	25	32	32	32	24	31	27	29	28	-	-
MSE + PLMI	0.77	81	3/6	32	12	16	2	32	-	-	-	-	-	-	-	2
MSE + DSS + PLMI	0.89	91	5/12	9	22	15	8	32	5	11	30	7	13	4	-	4

commercial sensor. The commercial sensor allowed the export of the recorded T_{cs} and H_{cs} measurements in the form of comma-separated value files via an Android application. The commercial sensor output contained measurement values at a sampling rate of 0.001667Hz (one sample per 10 minutes). Hence, we retrospectively processed the T_{RPi} and H_{RPi} measurements captured by the DHT22 sensor to match the number of samples and the measurement timestamps corresponding to the commercial sensor via the following procedure. For every timestamp (ts_{cs}) at which we had the temperature and humidity values from the commercial sensor, we constructed a 10-minute window which spanned from $ts_{cs} - (10 \text{ minutes})$ to ts_{cs} . We collated all T_{RPi} and H_{RPi} measurements in this time window and computed the mean value of these measurements to obtain $T_{RPi-\mu}[ts_{cs}]$ and $H_{RPi-\mu}[ts_{cs}]$. We then compared the $T_{RPi-\mu}$ with T_{cs} and $H_{RPi-\mu}$ with H_{cs} by plotting the signals one over the other. Further, we performed correlation analysis and fit separate linear models for the temperature and humidity measurements. Finally, we created separate Bland-Altman plots for the temperature and humidity measurements.

4.4 Results

4.4.1 Obstructive sleep apnea classification using a passive infrared sensor

Results obtained for the six experiments are presented in Table 4.4. The numbers under each feature indicated the frequency (out of the 32 LOOCV folds) with which

Table 4.5: Note Classification in Medical Equipment Alarm

Setting	$\mathcal{F}1 - micro$	$\mathcal{F}1 - macro$
Without Speech	0.98	0.97
With Speech Mixing	0.93	0.91

they were chosen for each fold. The bolded frequencies indicate that forward selection resulted in the inclusion of the feature. The highest \mathcal{F}_1 of 0.89 and Acc of 91% (bolded in columns two and three) were obtained by using five of the 12 features (when the feature combination MSE + DSS + PLMI were used). The top five features consistently selected across all folds associated with the model that provided the highest \mathcal{F}_1 and Acc were (in order of frequency of selection) \mathcal{H}_5 , σ_d^2 , \mathcal{H}_2 , \mathcal{H}_3 and \mathcal{K}_d . Further, without PLMI (the only non-PIR sensor based feature) a highest \mathcal{F}_1 of 0.83 and an Acc of 84% were obtained by using 10 features. Notably, the least frequently selected feature was the mean time between movements, μ_d .

Utilizing MSE alone, a highest \mathcal{F}_1 of 0.81 and Acc of 84% was obtained using three of the five MSE features. Usage of the DSS features alone, resulted in a highest \mathcal{F}_1 and Acc of 0.62 and 66% respectively whereas, usage of PLMI feature alone resulted in a \mathcal{F}_1 of 0.29 and Acc of 53% with the former experiment needing four of six features to attain the highest performance metrics.

4.4.2 Medical equipment alarm classification using audio

Table Table 4.5 shows the results for the 11-class medical equipment alarm note classification. We computed the micro averaged ($\mathcal{F}1 - micro$) and the macro-averaged ($\mathcal{F}1 - macro$) $\mathcal{F}1$ scores for the two experiments. When there was no speech content in the alarm audio recordings, we obtained an $\mathcal{F}1 - micro$ equal to 0.98 and an $\mathcal{F}1 - macro$ equal to 0.97. Retraining with speech resulted in a drop of 5.1% and 6.2% in $\mathcal{F}1 - micro$ and $\mathcal{F}1 - macro$ scores, respectively.

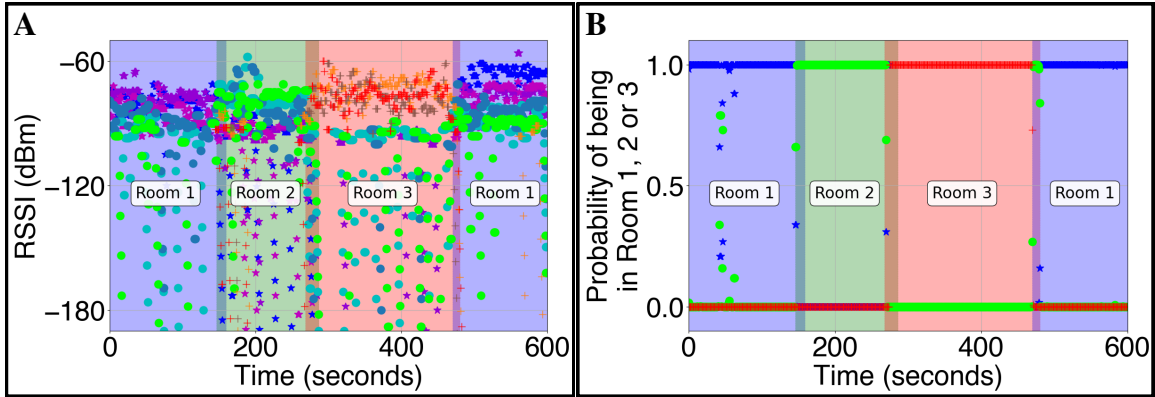


Figure 4.2 Human geolocation via Bluetooth. The translucent background colors indicate the ground truth values, and the corresponding room number has been labeled. (A) The received signal strength indicator (RSSI) values as captured by each of the nine Raspberry Pis during the experiment are shown. It is measured in decibels with reference to one milliwatt (dBm). (B) The corresponding probability values of the subject being in Rooms 1, 2, or 3 during the experiment. The blue color depicts Room 1, the green color depicts Room 2, and the red color depicts Room 3.

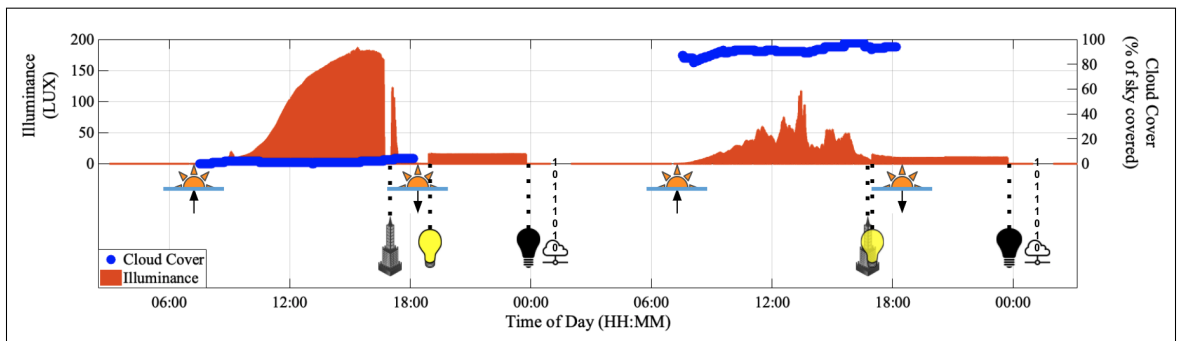


Figure 4.3 Observing the effects of cloud cover, sunrise, sunset, lights-ON, lights-OFF and buildings around the data collection site on the ambient light in a built environment for two continuous days (48 hours). The solid-orange bars (■) depict the amount of ambient light sensed by the color sensor in Lumen per square meter (LUX). The solid-blue circles (●) depict the local cloud cover in percentage of local sky covered by clouds. The sunrise and sunset times are indicated with the rising and setting sun symbols, respectively, using the upward and downward arrows. The yellow and black bulbs specify the lights-ON and lights-OFF times, respectively. The skyscraper symbol indicates the time when the Sun goes behind a skyscraper and causes a shadow onto the location where ambient light was being tracked. Data upload is depicted by binary values and a cloud node.

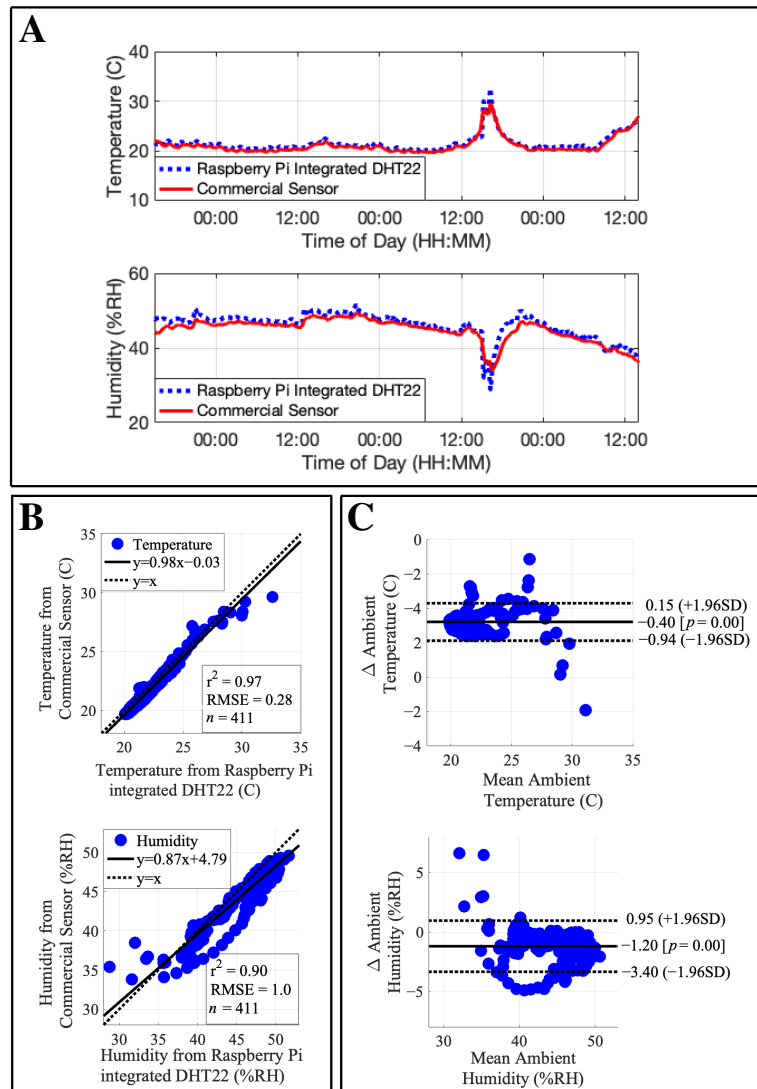


Figure 4.4 Comparison of temperature and humidity values captured by the Raspberry Pi (RPI) integrated DHT22 sensor and a commercial sensor. The top plot corresponds to temperature in each subplot, and the bottom plot corresponds to humidity values. (A) The dashed blue lines indicate the processed temperature and humidity values captured by the RPI integrated DHT22 sensor. The solid red lines indicate the corresponding values captured by the commercial sensor. (B) The correlation plots between values captured from the DHT22 sensor and the commercial sensor. The solid-blue circles (\bullet) indicate individual temperature and humidity tuples. The linear fits on the data and their deviations from the 45° line are depicted in the plot. (C) The Bland-Altman plot between the two measurements. The solid black lines indicate the mean difference between the two measurements. The dashed black lines indicate the +1.96 and the -1.96 standard deviation (SD) lines for the difference between the two measurements.

4.4.3 Geolocation of humans in a built environment via Bluetooth

The geolocation computation using the RSSI data was performed on a central server. Figure Figure 4.2A shows interpolated RSSI values captured by each of the nine RPIs that were placed in the built environment, and Figure Figure 4.2B illustrates the corresponding probability of the subject being in rooms 1, 2, or 3. The ground truth of the subject's presence is shown using translucent colors in the background. Specifically, translucent blue denoted being present in room 1, translucent green denoted being present in room 2, and translucent red denoted being present in room 3. Further, the transition from one room to another was illustrated by overlapping colors corresponding to the two rooms. It is evident from Figure Figure 4.2 that our system did an excellent job of identifying the room in which the person was present. Specifically, for 592 out of 600 seconds, the human tracking system correctly identified the subject's presence in one of the three rooms, which corresponded to an accuracy of 98.67%.

4.4.4 Ambient light logging

Figure Figure 4.3 depicts the variation of ambient light intensity over two days. The minimum, median, and maximum illuminance values during the D_{clear} period were equal to 0.56 LUX, 71.08 LUX, and 186.92 LUX, respectively, whereas the minimum, median, and maximum illuminance values during the D_{cloudy} period were equal to 0.26 LUX, 26.93 LUX, and 117.41 LUX, respectively. Thus, the median difference in illuminance between the clear and cloudy days was equal to 44.15 LUX. The minimum, median and maximum illuminance between the lights-ON and lights-OFF times was equal to 15.08 LUX, 15.30 LUX, 16.66 LUX on the first night (N_1) and equal to 9.39 LUX, 10.15 LUX, 15.02 LUX on the second night (N_2). The illuminance was consistently equal to zero between the lights-OFF and sunrise times. Further, we observed a dip in illuminance when the Sun hid behind a

skyscraper and cast a shadow on the data collection site during the day.

4.4.5 Ambient temperature and humidity logging

Figure Figure 4.4 illustrates the comparison between the processed temperature and humidity values from the DHT22 sensor with the outputs from a commercial sensor. The temperature values from the two sensors closely followed each other with a root mean squared error (RMSE) between the two measurements equal to 0.28°C and a coefficient of determination (r^2) equal to 0.97. Over 97% of the samples lay within the limits of agreement $[-1.96\text{SD}, +1.96\text{SD}]$ in the Bland-Altman plot. Further, the mean difference was equal to -0.4°C . The humidity values from the two sensors closely followed each other with an RMSE between the two measurements equal to $1.00\%\text{RH}$ and an r^2 equal to 0.90. Over 95% of the samples lay within the limits of agreement $[-1.96\text{SD}, +1.96\text{SD}]$ in the Bland-Altman plot. Further, the mean difference was equal to $-1.2\%\text{RH}$.

4.5 Discussion and Conclusion

The work described in this chapter aims to extend the types of data found in traditional clinical monitoring environments and provide a simple system to capture data in the built environment, outside of clinical settings. Many commercial (clinical and consumer) systems are either designed to keep data in a proprietary ‘walled-garden’ to reduce competition or are not designed for the high throughput needed to transmit/record the data. The RPi-based edge computing system described in this work allows direct data import via USB and upload to the cloud asynchronously to overcome these issues.

We have included methods to capture audio data, physical movement, and location of subjects. As we have demonstrated, audio data allows capturing of all alarms in the clinical space. While some monitors transmit some of these events

or signals over the network, it is often costly or impossible to gain access to such data, and data integration and synchronization are highly problematic. Moreover, such systems do not provide a holistic picture of the environment. For example, the volume of the alarm, together with the background noise, contributes to noise pollution and has been shown to affect caregivers and patients alike [134, 135, 173]. Beyond alarms in the clinical environment, it is possible to capture whether a patient is being mechanically ventilated (and at what frequency), groans and expression of pain, and other non-verbal utterances. It is possible to add speaker and voice recognition to the code base, to identify who is speaking and about what, providing insight into clinical (and non-clinical) discussions that may provide additional diagnostic power. For instance, by differentiating patients from family members, it is possible to assess both the level of clinical team support and frequency of bedside visits and the social support that a patient may have (inferred by the number and duration of visitations by friends and family). Tracking the time clinicians spend with patients and the level of expertise available could help identify gaps in care. In addition, by tracking Bluetooth transmitter strength of body-worn devices (e.g., badges or phones), it is possible to infer motion, an individual's identity (through a look-up table), and even the exact location of an individual if more than one Bluetooth receiver base is used. Real-time and accurate tracking of humans using Bluetooth beacons needs a receiver (RPi) sensor network. Further, we can have a central server where all the RPi's communicate and update the collected RSSI values. We can then have algorithms operating on this database in real-time to perform the geolocation of humans. We have implemented this system in a clinical environment at Emory Healthcare, Atlanta, USA, to monitor the real-time location of humans.

By capturing motion via video, we can probe even deeper into assessing the patient and their environment. For instance, we can estimate the quantity of sleep, sedation, and agitation that a patient experiences, all of which have been linked

to recovery [174]. At the same time, if the motion is associated with clinical care, then the intensity of activity can indicate when treatments, observations, or specific activities (such as replacing drips) took place. While we know that the ratio of nursing staff to patient impacts outcomes [137], there are no studies that examine the time at the bedside and the actions taken at the bedside in terms of their impact on the outcome. However, it is known that time spent at the bedside is linked to improved patient satisfaction [175].

Finally, the data modalities we capture provide us a unique opportunity to perform multimodal analysis of the patient state. For instance, consider the case of human sleep monitoring in a home environment. All we need to do is to place the Bluetooth beacon in the patient's clothes before they sleep. The motion signals captured during the patient's sleep give us the timestamps when the patient moved in the bed. Based on the intensity of the motion signal, we can delineate minor movements (rolling over) from major movements (sitting up in bed). The simultaneous recordings of audio-features and illuminance, which can act both as sleep inhibitors (flushing toilet or turning ON lights) and wake-event markers, give us valuable information about the patient's sleeping patterns. Further, the RSSI signal analysis will provide us with all the times that the patient exits the room during the night. Finally, via a long-term monitoring protocol, we can recommend ideal sleeping conditions to the patient by monitoring the temperature and humidity in the room. Thus, the system presented here provides a low-cost method for performing deep analysis, both at home and in a clinical setting.

CHAPTER 5

COMBINING WEARABLES AND NEARABLES FOR SLEEP-WAKE DETECTION

5.1 Introduction

In this chapter, we explored combining data from a wearable and a nearable for sleep-wake classification. Specifically, we used the Empatica E4 as the wearable and PSG-Video recorder as the nearable to monitor the sleep of a mixed-disorder elderly male population. We recorded PPG and three axis-acceleration from the wearable and extracted patient privacy-preserving body movement signals from the video recordings. Further, we extracted features from each signal modality and fed them to a binary classifier to perform sleep-wake detection. The goal of the study was to understand the effectiveness of video based features for sleep-wake detection. Further, we wanted to understand the benefit of combining wearable and nearable data modalities in sleep-wake detection.

5.2 Materials and methods

5.2.1 Dataset

We used a subgroup of participants included in the Emory Twin Study Follow-up and recruited from the Vietnam Era Twin Registry [164]. Overnight in-lab PSGs (including video) were recorded from 79 participants along with the PPG and three-axis acceleration from the wrist worn Empatica device. All data were collected at the Emory sleep-labs and the sleep staging was performed by in-house expert scorers at this facility. We obtained written informed consent from all participants, and the Emory University institutional review board approved this research

(IRB#00081004). Data were collected between March 2017 and December 2018 on the 79 male participants (aged $\in [61, 73]$ at the time of data collection). The cohort consisted of participants with mixed-disorders. Specifically, 23 participants suffered from mild ($5 < \text{AHI} \leq 15$), 7 participants from moderate ($15 < \text{AHI} \leq 30$) and 20 participants from severe ($\text{AHI} > 30$) sleep apnea. Further, 15 participants had mild ($5 < \text{PLMI} \leq 25$), 16 participants had moderate ($25 < \text{PLMI} \leq 50$) and 15 participants had severe ($\text{PLMI} > 50$) periodic leg movement disorder. Finally, 33 of the 79 participants in the study were obese (Body mass index > 30). The labels were generated by experienced sleep scorers at the Emory sleep labs who labeled each 30 second epoch as one of the following 5 categories: (1) Wake, (2) REM sleep (3) Sleep stage N1, (4) Sleep stage N2, and (5) Sleep stage N3. In our work, we combined categories two to four (REM sleep and sleep stages N1, N2 and N3) to obtain the *sleep* class. Thus, we tackled the binary classification problem of sleep-wake detection.

5.2.2 Preprocessing and feature extraction

The Empatica E4 captured three-axis acceleration signals at a sampling frequency equal to 32 Hz. We cleaned the signal via an equiripple FIR band-pass filter with the lower and higher pass band frequencies equal to 0.3 Hz and 11 Hz, and the lower and higher stop band frequencies equal to 0.25 Hz and 11.05 Hz. Next, we used the Oakley algorithm [176] to compute activity count (E) signals for each of x , y and z -axis acceleration signals (a_x, a_y, a_z) and for the root mean squared acceleration signal ($a_{rms} = \sqrt{a_x^2 + a_y^2 + a_z^2}$). For a given 30-second epoch $E[i]$, we computed the following two sleep-related metrics from each of the four above described activity count signals:

(1) Oakley sleep metric ($A[i]$) [176, 177] given by,

$$A[i] = 0.04 \times E[i - 2] + 0.2 \times E[i - 1] + 2 \times E[i] + 0.2 \times E[i + 1] + 0.04 \times E[i + 2] \quad (5.1)$$

(2) Kosmadopoulos sleep metric ($B[i]$) [178] given by,

$$B[i] = 0.04 \times E[i - 4] + 0.04 \times E[i - 3] + 0.2 \times E[i - 2] + 0.2 \times E[i - 1] \\ + 2 \times E[i] + 0.2 \times E[i + 1] + 0.2 \times E[i + 2] + 0.04 \times E[i + 3] + 0.04 \times E[i + 4] \quad (5.2)$$

We also computed the tilt-angle signals (ρ_x , ρ_y and ρ_z) for all three axes by extending the tilt-angle signal equation described by Cakmak et al. [93] as follows:

$$\rho_x = \frac{a_x}{\sqrt{a_x^2 + a_y^2 + a_z^2}} \\ \rho_y = \frac{a_y}{\sqrt{a_x^2 + a_y^2 + a_z^2}} \\ \rho_z = \frac{a_z}{\sqrt{a_x^2 + a_y^2 + a_z^2}} \quad (5.3)$$

In the end, we had the following signals:

- Activity counts: E_x , E_y , E_z and E_{rms}
- Oakley sleep metrics: A_x , A_y , A_z and A_{rms}
- Kosmadopoulos sleep metrics: B_x , B_y , B_z and B_{rms}
- Tilt-angle time series: ρ_x , ρ_y and ρ_z

The PPG signal was captured at a sampling frequency equal to 64 Hz. First, a simple baseline wander filtering was performed as described by Khaustov et al. [179]. Further, an equiripple FIR low-pass filter with a pass band frequency

= 5 Hz and a stop band frequency = 8 Hz was used to remove high frequency noise in the PPG signal. Finally, the beat-to-beat interval (NN) signal was extracted from the filtered PPG signal utilizing the *qppg* algorithm provided by Vest et al. [180]. The *qppg* algorithm measured the onset position of the PPG beats and was optimized to work well on PPG signals derived from adult humans. After NN-signal extraction, the respiratory induced amplitude variation (RIAV) signal was extracted as the difference between the maximum value of the PPG amplitude within a 75ms window from the PPG onset and the signal amplitude at a given PPG onset. Both the NN-signal and the RIAV-signal were resampled to 4 Hz for cardiorespiratory coupling (CRC) spectrogram calculations. The CRC spectrograms were computed as described by Li et al. [181, 94]. In summary, for each 30-second epoch of sleep or wake, we considered the NN-signal snippet and RIAV-signal snippet of length 5-minutes centered at the 30-second epoch. We then computed the CRC spectrogram by using a window of length 128 seconds with a shift of 10 seconds. The computed CRC spectrogram acted as the input for the pre-trained sleep-wake detector model and for the transfer learned sleep-wake detector model described in subsection 5.2.3.

The PSG-videos at Emory sleep labs were captured at a sampling rate equal to 5 Hz. The pixel resolution ($M \times N$) was 240×180 , $M = 240$ was the width and $N = 180$ was the height of a video frame within the video. To obtain body movement signals from these videos, we used two methods listed below:

- 1 Global difference mean (*GDM*) on the raw difference frames.
- 2 Scaled global difference mean on the Gaussian blurred and Otsu thresholded difference frames (GDM_{Otsu}).

The computation of *GDM* has been described in detail in subsection 4.2.1 except for the scaling-factor. Please refer to Equation 4.1 for the computation of *GDS*. The

additional step in computing the GDM was to scale the GDS signal by the total number of pixels in a frame ($M \times N$). Thus, we used a scaling-factor = $240 \times 180 = 43200$ to convert GDS to GDM signals. The GDM_{Otsu} signal was obtained by performing three additional intermediate steps when compared to the computation of GDS . First, an image smoothing step was performed on each difference frame D_t using a Gaussian blurring procedure to obtain D_t^G . The kernel size for the Gaussian window was set to 5×5 . The Gaussian blurring procedure helped in attenuating high frequency noise and assisted in improving body movement detection and suppressed spurious difference pixels. Next, we performed Otsu's thresholding method on D_t^G (grayscale image) so that, all pixels lower than the Otsu's threshold were set to zero. The Otsu's method determined the threshold in an unsupervised manner by accepting that the picture contained two classes of pixels (foreground and background), and such that the threshold maximized the separability of the resultant classes in gray levels. The difference frame obtained after Otsu's thresholding was denoted as D_t^{GO} . Then, the corresponding sample in the GDM_{Otsu} signal was obtained as:

$$GDM_{Otsu}[t] = \frac{\sum_{i \in [1, M]} \sum_{j \in [1, N]} D_t^{GO}}{\#\{\{d \in D_t^{GO} \mid |d| > \delta\}\}} \quad (5.4)$$

where the numerator summed the pixels that remained after Otsu's thresholding and the denominator was the count of non-zero pixels (scaling-factor) in the $M \times N$ frame. The PSG-videos captured contained a periodic noise with a period of 8 seconds. Both the GDM and GDM_{Otsu} signals exhibited this noise in the form of spikes at a frequency = $\frac{1}{8} = 0.125$ Hz. Thus, we performed feature extractions in the frequency domain rather than the time domain. For a given 30-second epoch (sleep or wake), we considered the GDM and GDM_{Otsu} signal snippets of length 5-minutes centered at the 30-second epoch. We computed the discrete-time Fourier

transform (DTFT) of both signal snippets using a Hanning window and a 4096-point fast Fourier transform. Next, we dropped the DTFT values at 0.125 Hz and its harmonics. We cropped the spectrum to only include values at frequencies $\in [0.0073, 0.5]$ Hz. This ensured that the mean value was suppressed from the body movement signals and only movements with a period of at least 2 seconds were considered in our analysis. Finally, we concatenated the remaining DTFT values and computed the mean, standard deviation, skewness and kurtosis for both the GDM and GDM_{Otsu} snippets to obtain eight dimensional feature vectors.

5.2.3 Experiments

We experimented with different feature combinations to understand the discriminatory power of different data modalities for sleep-wake detection. Specifically, we compared the classification performance of the below listed feature combinations.

- (1) Acc (Oakley): Features derived from three-axis acceleration
- (2) PPG (pre-trained): CRC derived from PPG with a pre-trained convolutional neural network (CNN) model
- (3) PPG (TL): CRC derived from PPG with a transfer learned CNN model
- (4) PPG (TL) + Acc (Oakley)
- (5) PPG + Acc (CPD): Change point detection based sleep-wake detector by Cakmak et al. [93]
- (6) Video-based movement (V): Features derived from GDM and GDM_{Otsu}
- (7) V + PPG (TL)
- (8) V + PPG (TL) + Acc (Oakley)

The first five experiments served as baseline models as they were based on previous works and the last three experiments showed us the improvements one can obtain via video-based analysis. The results for the change point detection based sleep wake detector (fifth in the list) have been directly included from the reference [93] without further validation. All other multimodal experiments were performed using the CatBoost classifier - A “Categorical Boosting” algorithm which has shown to outperform [182] the existing state-of-the-art implementations of gradient boosted decision trees such as XGBoost and LightGBM. Finally, each experiment (apart from experiment 5) were performed as is and after fixing imbalance in the training dataset by weighting the loss function by the proportion of samples in each class for either classes during training. All the models were trained and tested with a 10-fold by subject cross-validation. For the CatBoost classifiers trained in our experiments, the number of iterations needed for training was tuned by splitting the training data into further train-validation splits and utilizing the validation set to maximize the Cohen’s kappa [183, 105] between the model outputs and the ground truth labels for sleep-wake epochs. The remaining hyperparameters for the CatBoost classifier were set to the default values. Classification accuracy (Acu), sleep accuracy (Sleep-Acu), wake accuracy (Wake-Acu), area under the receiver operating curve (AUC) and Cohen’s Kappa (κ) were used to evaluate the performance of all the models. Cohen’s Kappa is calculated as

$$\kappa = \frac{p_a - p_e}{1 - p_e} \quad (5.5)$$

where $p_a = \sum_{k=0}^1 p_{kk}$, $p_e = \sum_{k=0}^1 p_{k+} p_{+k}$, p_{kk} represents the percentage of epochs classified into category k by the algorithm and by the annotated label; p_{k+} and p_{+k} represent the percentage of epochs assigned to category k by the algorithm and annotated label respectively.

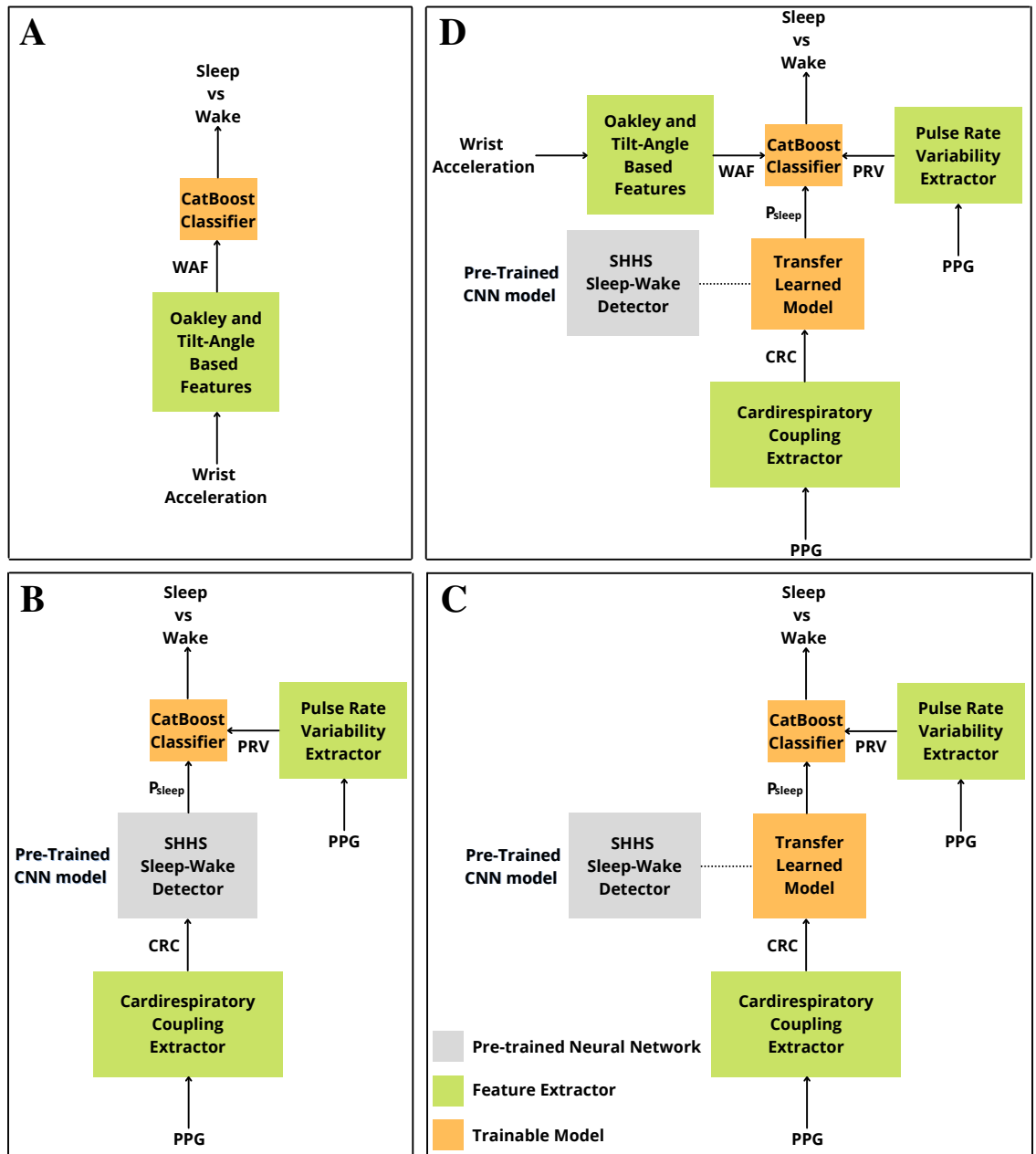


Figure 5.1 Baseline sleep wake detectors. (PPG: Photoplethysmogram; CRC: Cardiorespiratory Coupling; CNN: Convolutional Neural Network; SHHS: Sleep Heart Health Study; PRV: Pulse Rate Variability; P_{sleep} : Probability of sleep derived from the CNN model.)

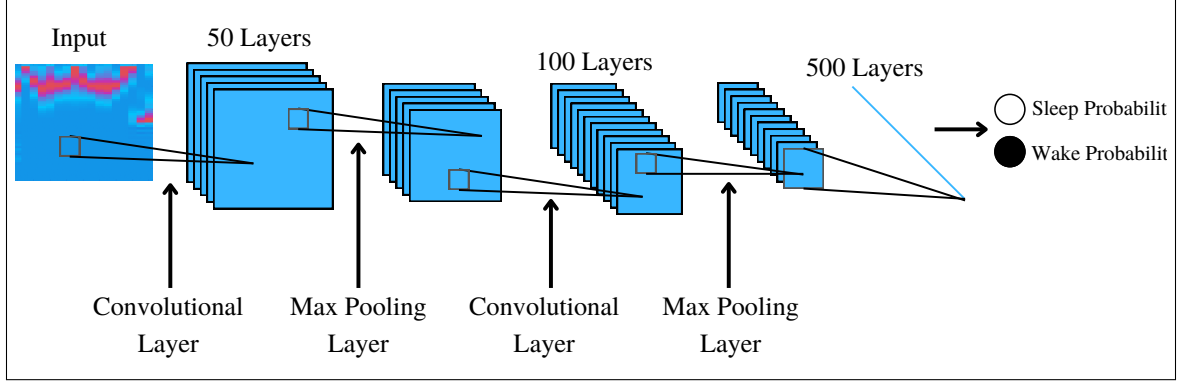


Figure 5.2 Convolutional neural network structure of the pre-trained model [181]. The input is a 50×18 cardiorespiratory coupling plot and the output layer contains two neurons representing the probabilities of wake (class ●) and sleep (class ○)

Sleep-wake detection from features derived from three-axis acceleration

We extracted a total of 18 features from the different signals derived from the three-axis acceleration. These 18 features included four activity count values - E_x , E_y , E_z and E_{rms} , four Oakley sleep metrics - A_x , A_y , A_z and A_{rms} , four Kosmadopoulos sleep metrics - B_x , B_y , B_z and B_{rms} , and six tilt-angle time series based features - $\mu(\rho_x)$, $\mu(\rho_y)$, $\mu(\rho_z)$, $\sigma(\rho_x)$, $\sigma(\rho_y)$ and $\sigma(\rho_z)$, where $\mu(\cdot)$ and $\sigma(\cdot)$ were the mean and standard deviation values computed over a 30-second window corresponding to the sleep-wake epoch. We trained a CatBoost classifier with these 18 features and pooled the results from a 10-fold by subject cross-validation. The top-left diagram in Figure 5.1 illustrates the three-axis accelerometer based sleep-wake detection.

Sleep-wake detection from PPG with a pre-trained CNN model

We derived a 50×18 CRC plot from the PPG signal for each sleep-wake epoch as described by Li et al. [181, 94]. The CRC plots were fed as inputs to a pre-trained CNN based sleep-wake detector model to obtain the probability of sleep (P_{sleep}). The architecture of the pre-trained CNN model is provided in Figure 5.2 and follows the same architecture that is described in [181]. The pre-training was performed

using the Sleep Heart Health Study visit 1 database [184] including 5793 overnight PSG recordings. One should note that the pre-trained CNN sleep-wake detection model was developed from a single lead ECG [181]. We derived 24 pulse rate variability (PRV) metrics [180] from the PPG signal for every sleep-wake epoch and combined them with P_{sleep} using a CatBoost classifier to obtain the final estimation of probability of sleep. The results were pooled from a 10-fold by subject cross-validation. The top-right diagram in Figure 5.1 illustrates the sleep-wake detection method from a pre-trained CNN model coupled with PRV metrics.

Sleep-wake detection from PPG with a transfer learned CNN model

In this experiment, we further trained the pre-trained model with the CRC plots derived from the Emory PPG database. In each training fold, we further trained the CNN model for 10 iterations over the entire training data and used early stopping [185] to determine the required number of iterations by computing the Cohen's kappa [183, 105] for a within training data validation set. Once the transfer learning of the CNN model was complete, we obtained the P_{sleep} for all of the training data and combined them with the corresponding PRV metrics to further train the CatBoost classifier. The sleep-wake detection performance was measured via the 10-fold by subject cross-validation. The bottom-left diagram in Figure 5.1 illustrates the sleep-wake detection method using CRC plots with a transfer learned CNN model coupled with PRV metrics.

Sleep-wake detection from combining features from accelerometry and PPG

In this experiment, we combined the features from experiments 1 and 3. The transfer learning procedure was replicated from experiment 3. The CatBoost classifier now received 18 three-axis acceleration based features, 24 PRV metrics and one P_{sleep} value from the transfer learned CNN. We measured the classification

performance using the same 10-fold by subject cross-validation that was used in the previous experiments. The bottom-right diagram in Figure 5.1 illustrates the accelerometry features, PRV metrics and CRC based sleep-wake detection.

Sleep-wake detection from video based movement

Until now, we looked at methods for sleep-wake detection using a wearable (the Empatica E4 watch). In this experiment, we tested using the features derived from signals captured using a video camera (a nearable) for sleep-wake detection. We derived eight dimensional feature vectors per epoch that included the following features: $\mu(GDM[i])$, $\sigma(GDM[i])$, $skew(GDM[i])$, $kurt(GDM[i])$, $\mu(GDM_{Otsu}[i])$, $\sigma(GDM_{Otsu}[i])$, $skew(GDM_{Otsu}[i])$ and $kurt(GDM_{Otsu}[i])$, where $\mu(\cdot)$, $\sigma(\cdot)$, $skew(\cdot)$ and $kurt(\cdot)$ were the statistical mean, standard deviation, skewness and kurtosis of the input signals and i was the epoch index which indicated the corresponding snippets of GDM and GDM_{Otsu} . We fed these video based body movement features to a CatBoost classifier and pooled the results from a 10-fold by subject cross-validation to measure the sleep-wake detection performance. The top-left diagram in Figure 5.3 illustrates the sleep-wake detection method using video based movement.

Sleep-wake detection from video based movement and PPG

Next, we combined the video based body movement features with the PRV metrics and the P_{sleep} value from the transfer learned CNN to perform sleep-wake detection. In total, the CatBoost classifier took 33 features as input for each epoch and performed binary classification. The results were pooled from a 10-fold by subject cross-validation. The top-right diagram in Figure 5.3 illustrates the sleep-wake detection method using video based movement, PRV metrics and CRC.

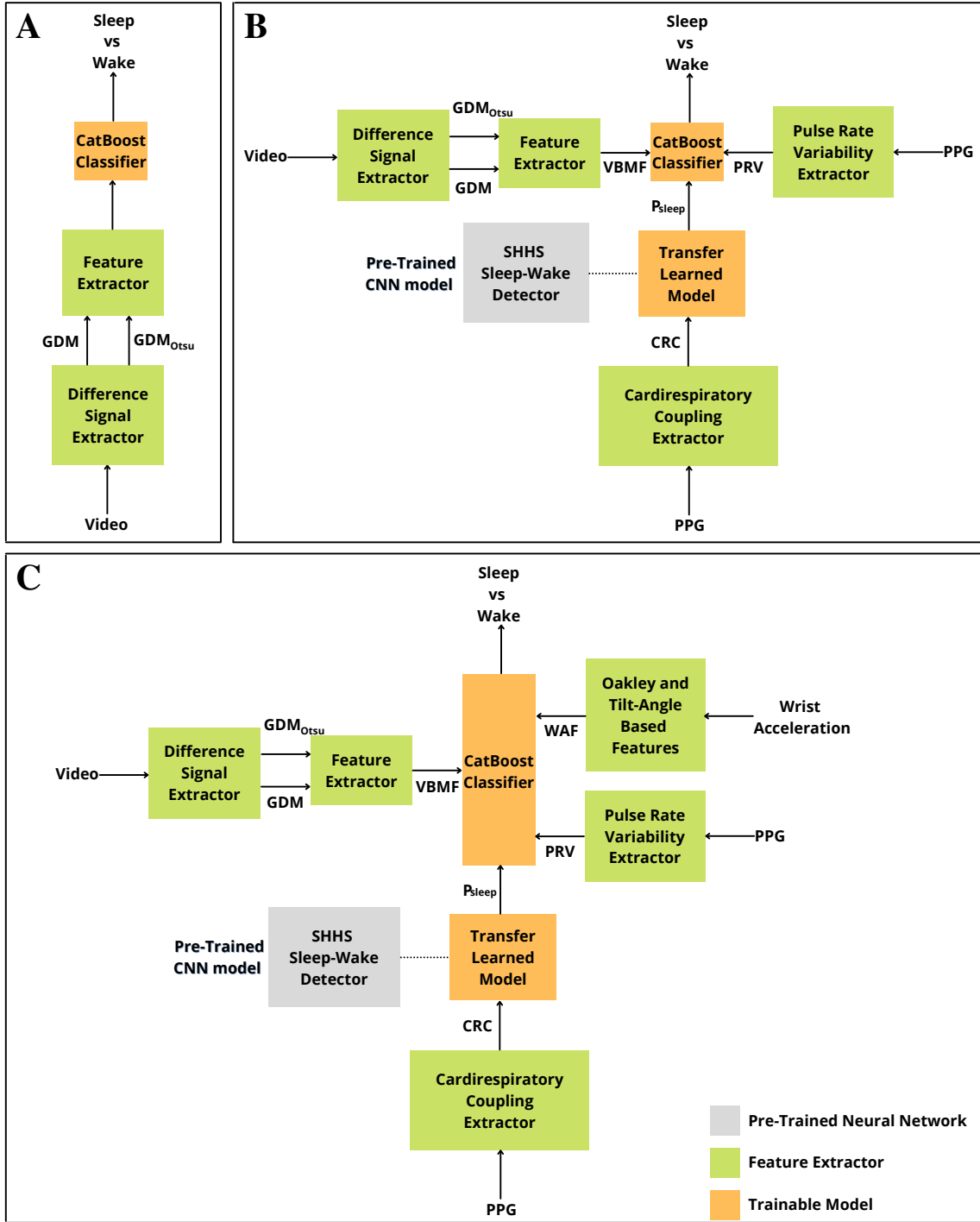


Figure 5.3 Proposed sleep wake detectors. (PPG: Photoplethysmogram; CRC: Cardiorespiratory Coupling; CNN: Convolutional Neural Network; SHHS: Sleep Heart Health Study; PRV: Pulse Rate Variability; P_{sleep} : Probability of sleep derived from the CNN model.)

Table 5.1: Performance of baseline sleep-wake detectors (Sleep-Acu = Sleep Accuracy, Wake-Acu = Wake Accuracy, κ = Kappa, SD = Standard deviation, Acc = Three-axis acceleration, PPG = Photoplethysmogram, TL = Transfer learned, CPD = Change point detection method)

Models	Balanced	Accuracy (%)		Sleep-Acu (%)		Wake-Acu (%)		F1-score (%)		κ	
		Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
Acc (Oakley)	No	71	(12)	86	(13)	41	(15)	79	(12)	0.28	(0.16)
	Yes	65	(10)	70	(18)	55	(18)	71	(14)	0.22	(0.15)
PPG (pre-trained)	No	71	(11)	81	(14)	53	(19)	78	(11)	0.32	(0.15)
	Yes	66	(12)	64	(19)	74	(16)	70	(15)	0.31	(0.15)
PPG (TL)	No	73	(10)	85	(12)	52	(18)	80	(10)	0.35	(0.14)
	Yes	70	(10)	71	(16)	70	(16)	75	(12)	0.36	(0.15)
PPG (TL) + Acc (Oakley)	No	75	(10)	86	(10)	54	(18)	81	(10)	0.38	(0.15)
	Yes	71	(9)	74	(15)	67	(17)	76	(11)	0.36	(0.13)
PPG + Acc (CPD) [93]	No	76	(9)	85	(12)	54	(20)	59	(14)	0.39	(0.17)

Table 5.2: Performance of proposed sleep-wake detectors (Sleep-Acu = Sleep Accuracy, Wake-Acu = Wake Accuracy, κ = Kappa, SD = Standard deviation, PPG = Photoplethysmogram, TL = Transfer learned, Acc = Three-axis acceleration)

Models	Balanced	Accuracy (%)		Sleep-Acu (%)		Wake-Acu (%)		F1-score (%)		κ	
		Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
Video features (V)	No	77	(10)	91	(5)	52	(15)	83	(9)	0.42	(0.13)
	Yes	74	(8)	79	(10)	67	(15)	79	(9)	0.41	(0.13)
V + PPG (TL)	No	78	(8)	89	(7)	59	(16)	83	(8)	0.45	(0.13)
	Yes	75	(8)	78	(12)	72	(16)	80	(9)	0.44	(0.13)
V + PPG (TL) + Acc (Oakley)	No	78	(8)	89	(7)	58	(17)	83	(8)	0.45	(0.13)
	Yes	75	(7)	79	(10)	69	(17)	80	(8)	0.43	(0.12)

Sleep-wake detection from combining wearables and nearables

Finally, we combined three-axis acceleration features along with video based body movement features, PRV metrics, and CRC plots. In this experiment, the CatBoost classifier took 51 features per epoch. The classification performance was measured by pooling the results from a 10-fold by subject based cross-validation. The bottom diagram in Figure 5.3 illustrates the sleep-wake detection method using video based movement, PRV metrics, CRC plots and three-axis acceleration features.

5.3 Results

Table Table 5.1 and Table 5.2 present the classification performance for all the baseline models and proposed models respectively. All the results provided (except

the CPD [93] model) were computed by computing the metrics for each participant and then taking the mean and standard deviation of the metrics for all participants.

The proposed models always outperformed their baseline counterparts indicating that the addition of video based features always improved sleep-wake detection. The F1-score and κ provide an all-round view (not biased to class imbalance) of the model performance and the imbalanced model with video based features and PPG based transfer learned CNN features performed the best in terms of F1-score (83%) and κ (0.45). The addition of features derived from three-axis acceleration did not improve its performance. The imbalanced model using just the video based features performed with the highest sleep accuracy (91%). The balanced PPG based pre-trained model had the highest wake accuracy (74%) amongst all the models. Among the baseline models, the CPD model [93] had the best accuracy (76%) and κ (0.39), whereas the imbalanced PPG (TL) + Acc (Oakley) model performed with the highest sleep accuracy (86%) and F1-score (81%).

5.4 Discussion and Conclusion

In this chapter, we presented a novel video based method for identifying sleep and wake states from body movements. The method comprised of extracting two difference signals: GDM and GDM_{Otsu} , and computing statistical features from these signals for a binary classification task. Our experiments showed the flexibility of the video based sleep-wake detector in that it could easily incorporate additional features from physiological signals (PPG) and wrist-worn accelerometer to improve classification performance. In the experiment where PPG was used in isolation, we showed that transfer learning helped increase the sleep-wake detection performance in terms of accuracy and κ . However, the wake accuracy did not improve with transfer learning. Combining the transfer learned model with video based features did improve the wake accuracy. The addition of features from the wrist-worn

accelerometer did not significantly improve the performance when compared to just using video and PPG derived features. Moreover, when comparing the model with just video features versus the model with just acceleration based (Oakley) features, the video model performed significantly better. This showed that the metrics derived from video based movement contained more discriminatory features for sleep-wake detection compared to wrist-worn acceleration. Since the video based body movement features capture global (entire body) movements and wrist-worn acceleration captures local (wrist and hand) movements, we conclude that capturing global body movements are more crucial than capturing the local wrist movements for sleep-wake detection. In conclusion, video based body movements are extremely useful in sleep-wake detection and perform significantly better than wrist-worn acceleration or PPG based sleep-wake detectors. Further, combining wearables and nearables does provide a boost in classification performance for sleep-wake detection compared to unimodal models.

CHAPTER 6

CONCLUSION

This dissertation presents wearable and nearable-based technologies for patient state analysis. The theme of this work is to keep the system low-cost and patient privacy-preserving while being able to monitor patient state at a high accuracy. All the methods presented in this thesis were validated using human-subject studies with data collected in the wild.

In chapter 3, we focused on developing a disease severity classifier for patients suffering from Rett syndrome using a wearable. First, we developed two techniques to reduce the amount of missing data in ECG signals. Next, we showed that HRV metrics combined with features capturing the interaction between the heart rate signal and body movements could accurately classify high severity Rett patients from low severity patients. Finally, we identified a physio-motor biomarker for Rett syndrome - *deceleration capacity of the heart rate during sleep* - it was a highly discriminatory feature for Rett syndrome severity detection with a feature popularity score equal to 1.

Although wearables have many advantages for patient monitoring, they have two primary disadvantages - localized measurements and missing data. To overcome these issues, in the next chapter, we developed techniques for patient monitoring with nearables. For this, we built a generalized open-source system for edge computing in clinical and home environments. It provided real-time data elements and analysis that were not generally present in electronic medical records yet were associated with clinical performance, diagnosis, and outcomes. In particular, we focused on the acoustic environment (such as speech, alarms, and environmental noise), human movement detection (from PIR sensor and IR camera) and

geolocation (absolute or relative to others). Further, we included methods for privacy-preserving feature extraction to provide a generally acceptable system that is unlikely to violate hospital policies and other privacy regulations, which may reduce the anxiety of administrators and clinicians concerning the level of monitoring. The implementation on a state-of-the-art extensible edge computing system at a relatively low cost provided a high degree of flexibility in the design. The bill-of-materials and open-source code to replicate the work described were made publicly available under an open-source license [156]. Using the PIR sensor based body movement detector, we showed it is possible to detect obstructive sleep apnea in a mixed-disorder elderly patient cohort at a high degree of accuracy. We also showed that one can classify medical equipment alarms using edge-computed features derived from audio. Further, we described an algorithm to accurately perform room-level geolocation of humans and to accurately track ambient environmental parameters (temperature, humidity and illuminance) in a built environment.

The last part of the dissertation presented a method to combine data from wearables and nearables for improved sleep-wake detection in a mixed-disorder elderly male cohort. We showed that video based analysis provided a boost in accuracy and κ values for sleep-wake detection when compared to wearable based analysis. Moreover, combining nearable data with wearable data further improved the classification performance. Most video based sleep-wake detectors have focused on infants [186, 187, 188] or have very small number of participants [96]. To the author's knowledge, this is the first large scale (> 50 patients) study in a mixed-disorder elderly male cohort to assess the effectiveness of features derived from video based body movements for sleep-wake detection.

6.1 Future work

The methods presented in this dissertation have performed at a high accuracy, however, we feel further studies are needed to stress-test the methods and improve generalizability. Thus, the future work includes applying the techniques presented in this thesis to larger and whenever possible in more diverse cohorts to fine-tune the algorithms as necessary for clinical usage.

6.1.1 Severity of Rett syndrome from wearables

The Rett syndrome severity estimation project included 20 patients who had continuous and simultaneous ECG and three-axis acceleration recorded for at least two days. The *Rett Syndrome Research Trust* (our collaborators in this project) have since collected the same wearable sensor data from an additional 20 patients who are suffering from Rett syndrome. The data imputation, feature extraction, classification and feature importance measurement techniques can be easily translated to these patients and thus, we can test the generalizability of the wearable data processing techniques described in this thesis.

6.1.2 Patient state analysis from nearables

The hardware system designed in this project has numerous potential applications. Specifically, these systems are very suited to monitor neurological populations as the data collection procedure is passive and does not rely on patients' or caregivers' compliance. Thus, the system itself has been deployed for patient monitoring and data collection in two separate healthcare facilities located in the United States of America: (1) A New York state department of health funded center for excellence facility that offers residential, medical, clinical and special education programs to the residents (25 units); and (2) A Mild Cognitive Impairment rehabilitation

program facility at Emory Healthcare in Atlanta (40 units). In terms of data volume, for a 12 hour (7 AM to 7 PM) recording of all data modalities, a total of 832 MB of data is recorded by the device. The data collection is ongoing and the processing of these datasets is part of future work.

6.1.3 Combining wearables and nearables for sleep-wake detection

One possible direction is to extend the current work of sleep-wake detection to both generalized sleep-staging and detecting sleep-disorders such as OSA and PLMD. Further, the algorithm needs to be stress-tested on larger databases and external cohorts to establish generalizability of the models. Finally, our method can be easily integrated into existing PSG analysis software to provide a first pass sleep-wake detection before experienced sleep scorers update the algorithm's estimate.

6.2 Final remarks

The combination of wearables and nearables have huge potential for patient state analysis. While both modalities have advantages and disadvantages, they compliment each other to provide a more complete understanding of the patient state. With the advent of newer paradigms in healthcare such as internet of things, multi-modal data analysis is the way forward. While missing data, time-synchronization of sensors and intelligent algorithms that can leverage multimodal data continue to be the grand challenges in the field, this dissertation provides concrete methods which address each of these concerns in specific settings. In conclusion, this dissertation provides a foundation for capturing and combining physiological data from wearables and passive global data from nearables for patient state analysis.

REFERENCES

- [1] R. Williams, "Neurology at a distance," *The Lancet Neurology*, vol. 9, no. 4, pp. 346–347, 2010.
- [2] S. Beladakere Ramaswamy, S. M. Bhagavan, and P. K. Sahota, "Remote monitoring in tele-neurology (sleep apnea case)," in *Learning Teleneurology Basics*, Springer, 2021, pp. 125–138.
- [3] J. Dunn, R. Runge, and M. Snyder, "Wearables and the medical revolution," *Personalized medicine*, vol. 15, no. 5, pp. 429–448, 2018.
- [4] J. Klucken, T. Gladow, J. G. Hilgert, M. Stamminger, C. Weigand, and B. Eskofier, "Wearables in the treatment of neurological diseases-where do we stand today?" *Der Nervenarzt*, vol. 90, no. 8, pp. 787–795, 2019.
- [5] MC10®Biostamp, <https://www.mc10inc.com/>, [Online; accessed 01-Apr-2021].
- [6] L. Piwek, D. A. Ellis, S. Andrews, and A. Joinson, "The rise of consumer health wearables: Promises and barriers," *PLOS Medicine*, vol. 13, e1001953, 2 Feb. 2016.
- [7] J. Dunn, R. Runge, and M. Snyder, "Wearables and the medical revolution," *Personalized Medicine*, vol. 15, pp. 429–448, 5 Sep. 2018.
- [8] A. K. Yetisen, J. L. Martinez-Hurtado, B. Ünal, A. Khademhosseini, and H. Butt, "Wearables in medicine," *Advanced Materials*, vol. 30, p. 1706910, 33 Aug. 2018.
- [9] M. Chen, S. Gonzalez, A. Vasilakos, H. Cao, and V. C. M. Leung, "Body area networks: A survey," *Mobile Networks and Applications*, vol. 16, pp. 171–193, 2 Apr. 2011.
- [10] P. A. Catherwood, S. S. Bukhari, G. Watt, W. G. Whittow, and J. McLaughlin, "Body-centric wireless hospital patient monitoring networks using body-contoured flexible antennas," *IET Microwaves, Antennas & Propagation*, vol. 12, pp. 203–210, 2 Feb. 2018.
- [11] S. Movassaghi, M. Abolhasan, J. Lipman, D. Smith, and A. Jamalipour, "Wireless body area networks: A survey," *IEEE Communications Surveys Tutorials*, vol. 16, no. 3, pp. 1658–1686, 2014.

- [12] E. Monton, J. Hernandez, J. Blasco, T. Herve, J. Micallef, I. Grech, A. Brincat, and V. Traver, "Body area network for wireless patient monitoring," *IET Communications*, vol. 2, p. 215, 2 2008.
- [13] S. LaFleur and I. Mahgoub, "A personal body area network as a pre-screening surrogate to the polysomnography," in *Proceedings of the 8th International Conference on Body Area Networks*, 2013, pp. 233–236.
- [14] B. Gyselinckx, J. Penders, and R. Vullers, "Potential and challenges of body area networks for cardiac monitoring," *Journal of Electrocardiology*, vol. 40, no. 6, Supplement 1, S165–S168, 2007.
- [15] L. R. Martin, S. L. Williams, K. B. Haskard, and M. R. DiMatteo, "The challenge of patient adherence," *Therapeutics and clinical risk management*, vol. 1, no. 3, p. 189, 2005.
- [16] E. Gederer and G. D. Clifford, "Fusion of image and signal processing for the detection of obstructive sleep apnea," in *Proceedings of 2012 IEEE-EMBS International Conference on Biomedical and Health Informatics*, IEEE, 2012, pp. 890–893.
- [17] A. S. Devlin and A. B. Arneill, "Health care environments and patient outcomes: A review of the literature," *Environment and behavior*, vol. 35, no. 5, pp. 665–694, 2003.
- [18] *Rett syndrome*, <https://www.rettssyndrome.org/>, [Online; accessed 01-Apr-2021].
- [19] D. C. Tarquinio and A. K. Percy, "Rett syndrome: Clinical aspects," in *Neuronal and Synaptic Dysfunction in Autism Spectrum Disorder and Intellectual Disability*, Elsevier, 2016, pp. 301–323.
- [20] D. G. Glaze, "Neurophysiology of rett syndrome," *Journal of child neurology*, vol. 20, no. 9, pp. 740–746, 2005.
- [21] E. A. Sekul, J. P. Moak, R. J. Schultz, D. G. Glaze, J. K. Dunn, and A. K. Percy, "Electrocardiographic findings in rett syndrome: An explanation for sudden death?" *The Journal of pediatrics*, vol. 125, no. 1, pp. 80–82, 1994.
- [22] M. Fuster-Siebert and M. Castro-Gago, "Electrocardiographic findings in rett syndrome," *The Journal of Pediatrics*, vol. 126, no. 3, p. 506, 1995.
- [23] C. Ellaway, G. Sholler, H. Leonard, and J. Christodoulou, "Prolonged qt interval in rett syndrome," *Archives of disease in childhood*, vol. 80, no. 5, pp. 470–472, 1999.

- [24] J. Singh, S. Ameenpur, R. Ahmed, S. Basheer, S. Chishti, R. Lawrence, F. Fiori, and P. Santosh, "An observational study of heart rate variability using wearable sensors provides a target for therapeutic monitoring of autonomic dysregulation in patients with rett syndrome," *Biomedicines*, vol. 10, no. 7, p. 1684, 2022.
- [25] P. O. O. Julu, A. M. Kerr, F. Apartopoulos, S. Al-Rawas, I. W. Engerström, L. Engerström, G. A. Jamal, and S. Hansen, "Characterisation of breathing and associated central autonomic dysfunction in the rett disorder," *Archives of disease in childhood*, vol. 85, pp. 29–37, 1 2001.
- [26] F. Guideri, M. Acampa, T. DiPerri, M. Zappella, and Y. Hayek, "Progressive cardiac dysautonomia observed in patients affected by classic rett syndrome and not in the preserved speech variant," *Journal of Child Neurology*, vol. 16, pp. 370–373, 5 2001.
- [27] F. Guideri, M. Acampa, P. Bardi, A. D. Lalla, M. Zappella, and Y. Hayek, "Cardiac dysautonomia and serotonin plasma levels in rett syndrome," *Neuropediatrics*, vol. 35, pp. 36–38, 01 2004.
- [28] A. Kumar, A. Jaryal, S. Gulati, B. Chakrabarty, A. Singh, K. K. Deepak, R. M. Pandey, N. Gupta, S. Sapra, M. Kabra, *et al.*, "Cardiovascular autonomic dysfunction in children and adolescents with rett syndrome," *Pediatric neurology*, vol. 70, pp. 61–66, 2017.
- [29] J. Singh and P. Santosh, "Key issues in rett syndrome: Emotional, behavioural and autonomic dysregulation (ebad)-a target for clinical trials," *Orphanet journal of rare diseases*, vol. 13, pp. 1–17, 1 2018.
- [30] D. C. Tarquinio, W. Hou, J. L. Neul, G. K. Berkmen, J. Drummond, E. Aronoff, J. Harris, J. B. Lane, W. E. Kaufmann, K. J. Motil, *et al.*, "The course of awake breathing disturbances across the lifespan in rett syndrome," *Brain and Development*, vol. 40, pp. 515–529, 7 2018.
- [31] F. Guideri and M. Acampa, "Sudden death and cardiac arrhythmias in rett syndrome," *Pediatric cardiology*, vol. 26, p. 111, 1 2005.
- [32] M. Acampa and F. Guideri, "Cardiac disease and rett syndrome," *Archives of Disease in Childhood*, vol. 91, pp. 440–443, 5 May 2006.
- [33] M. S. Carroll, J.-M. Ramirez, and D. E. Weese-Mayer, "Diurnal variation in autonomic regulation among patients with genotyped rett syndrome," *Journal of Medical Genetics*, vol. 57, pp. 786–793, 11 Nov. 2020.

- [34] D. E. Weese-Mayer, S. P. Lieske, C. M. Boothby, A. S. Kenny, H. L. Bennett, J. M. Silvestri, and J.-M. Ramirez, "Autonomic nervous system dysregulation: Breathing and heart rate perturbation during wakefulness in young girls with rett syndrome," *Pediatric research*, vol. 60, pp. 443–449, 4 2006.
- [35] D. E. Weese-Mayer, S. P. Lieske, C. M. Boothby, A. S. Kenny, H. L. Bennett, and J.-M. Ramirez, "Autonomic dysregulation in young girls with rett syndrome during nighttime in-home recordings," *Pediatric pulmonology*, vol. 43, pp. 1045–1060, 11 2008.
- [36] P. M. FitzGerald, J. Jankovic, and A. K. Percy, "Rett syndrome and associated movement disorders," *Movement disorders: official journal of the Movement Disorder Society*, vol. 5, pp. 195–202, 3 1990.
- [37] E. Roze, V. Cochen, S. Sangla, T. Bienvenu, A. Roubergue, S. Leu-Semenescu, and M. Vidaihet, "Rett syndrome: An overlooked diagnosis in women with stereotypic hand movements, psychomotor retardation, parkinsonism, and dystonia?" *Movement disorders: official journal of the Movement Disorder Society*, vol. 22, pp. 387–389, 3 2007.
- [38] T. Temudo, P. Maciel, and J. Sequeiros, "Abnormal movements in rett syndrome are present before the regression period: A case study," *Movement disorders: official journal of the Movement Disorder Society*, vol. 22, pp. 2284–2287, 15 2007.
- [39] T. Temudo, E. Ramos, K. Dias, C. Barbot, J. P. Vieira, A. Moreira, E. Calado, I. Carrilho, G. Oliveira, A. Levy, *et al.*, "Movement disorders in rett syndrome: An analysis of 60 patients with detected mecp2 mutation and correlation with mutation type," *Movement disorders: official journal of the Movement Disorder Society*, vol. 23, pp. 1384–1390, 10 2008.
- [40] N. R. Oakley, "Validation with polysomnography of the sleepwatch sleep/wake scoring algorithm used by the actiwatch activity monitoring system," *Bend: Mini Mitter, Cambridge Neurotechnology*, 1997.
- [41] A. S. Cakmak, E. Reinertsen, H. A. Taylor, A. J. Shah, and G. D. Clifford, "Personalized heart failure severity estimates using passive smartphone data," *IEEE*, Dec. 2018, pp. 1569–1574, ISBN: 978-1-5386-5035-6.
- [42] O. Hallioglu, C. Okuyaz, E. Mert, and K. Makharoblidze, "Effects of antiepileptic drug therapy on heart rate variability in children with epilepsy," *Epilepsy research*, vol. 79, no. 1, pp. 49–54, 2008.

- [43] P. A. Lotufo, L. Valiengo, I. M. Bensenor, and A. R. Brunoni, "A systematic review and meta-analysis of heart rate variability in epilepsy and antiepileptic drugs," *Epilepsia*, vol. 53, no. 2, pp. 272–282, 2012.
- [44] H. XING, Y. SHEN, H. CHEN, Y. WANG, and W. SHEN, "Heart rate variability and its response to thyroxine replacement therapy in patients with hypothyroidism," *Chinese Medical Journal*, vol. 114, no. 09, pp. 906–908, 2001.
- [45] E. L. Jones, S. Perring, A. Khattab, and O. Allenby-Smith, "The effects of proton pump inhibitors on autonomic tone in patients with erosive and non-erosive esophagitis," *Neurogastroenterology & Motility*, vol. 28, no. 5, pp. 659–664, 2016.
- [46] F. Cottin, V. Malcurat, H. Zorgati, F. Prieur, Z. Labsy, M. Do, O. Gagey, and K. Collomp, "Effect of oral glucocorticoid intake on autonomic cardiovascular control," *SpringerPlus*, vol. 4, no. 1, pp. 1–9, 2015.
- [47] F. Lin, Y. Zhuang, C. Song, A. Wang, Y. Li, C. Gu, C. Li, and W. Xu, "Sleepsense: A noncontact and cost-effective sleep monitoring system," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 11, pp. 189–202, 1 Feb. 2017.
- [48] Q. Liang, L. Xu, N. Bao, L. Qi, J. Shi, Y. Yang, and Y. Yao, "Research on non-contact monitoring system for human physiological signal and body movement," *Biosensors*, vol. 9, p. 58, 2 Apr. 2019.
- [49] F. Wang, M. Skubic, M. Rantz, and P. E. Cuddihy, "Quantitative gait measurement with pulse-doppler radar for passive in-home gait assessment," *IEEE Transactions on Biomedical Engineering*, vol. 61, pp. 2434–2443, 9 Sep. 2014.
- [50] Y. S. Lee, P. N. Pathirana, T. Caelli, and R. Evans, "Doppler radar in respiratory monitoring: Detection and analysis," *IEEE*, Nov. 2013, pp. 224–228, ISBN: 978-1-4799-0572-0.
- [51] C. Gu, C. Li, J. Lin, J. Long, J. Huangfu, and L. Ran, "Instrument-based non-contact doppler radar vital sign detection system using heterodyne digital quadrature demodulation architecture," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, pp. 1580–1588, 6 Jun. 2010.
- [52] D. Drzymalski, J. Ceruzzi, and W. Camann, "Noise in the obstetric operating room," *International Journal of Obstetric Anesthesia*, vol. 29, pp. 87–88, 2017.
- [53] S. Keller, F. Tschan, N. K. Semmer, E. Holzer, D. Candinas, M. Brink, and G. Beldi, "Noise in the operating room distracts members of the surgical team.

- An observational study," *World Journal of Surgery*, vol. 42, no. 12, pp. 3880–3887, 2018.
- [54] V. Murthy, S. Malhotra, I. Bala, and M. Raghunathan, "Detrimental effects of noise on anaesthetists," *Canadian Journal of Anaesthesia*, vol. 42, no. 7, p. 608, 1995.
- [55] R. A. Shapiro and T. Berland, "Noise in the operating room," *New England Journal of Medicine*, vol. 287, no. 24, pp. 1236–1238, 1972.
- [56] T. J. Way, A. Long, J. Weihing, R. Ritchie, R. Jones, M. Bush, and J. B. Shinn, "Effect of noise on auditory processing in the operating room," *Journal of the American College of Surgeons*, vol. 216, no. 5, pp. 933–938, 2013.
- [57] S. Guerlain, R. B. Adams, F. B. Turrentine, T. Shin, H. Guo, S. R. Collins, and J. F. Calland, "Assessing team performance in the operating room: Development and use of a "black-box" recorder and other tools for the intraoperative environment," *Journal of the American College of Surgeons*, vol. 200, no. 1, pp. 29–37, 2005.
- [58] C. Azevedo-Coste, R. Pissard-Gibollet, G. Toupet, É. Fleury, J.-C. Lucet, and G. Birgand, "Tracking clinical staff behaviors in an operating room," *Sensors*, vol. 19, no. 10, p. 2287, 2019.
- [59] F. Adib, Z. Kabelac, and D. Katabi, "Multi-person localization via RF body reflections," in *12th USENIX Symposium on Networked Systems Design and Implementation (NSDI 15)*, 2015, pp. 279–292.
- [60] S. Kianoush, S. Savazzi, F. Vicentini, V. Rampa, and M. Giussani, "Device-free RF human body fall detection and localization in industrial workplaces," *IEEE Internet of Things Journal*, vol. 4, no. 2, pp. 351–362, 2016.
- [61] Y. Tian, G.-H. Lee, H. He, C.-Y. Hsu, and D. Katabi, "RF-based fall monitoring using convolutional neural networks," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 3, pp. 1–24, 2018.
- [62] J. Aschoff, "Circadian rhythms: Influences of internal and external factors on the period measured in constant conditions," *Zeitschrift für Tierpsychologie*, vol. 49, no. 3, pp. 225–249, 1979.
- [63] S. S. Campbell and D. Dawson, "Enhancement of nighttime alertness and performance with bright ambient light," *Physiology & Behavior*, vol. 48, no. 2, pp. 317–320, 1990.

- [64] K. Honma, S. Honma, M. Kohsaka, and N. Fukuda, "Seasonal variation in the human circadian rhythm: Dissociation between sleep and temperature rhythm," *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, vol. 262, no. 5, R885–R891, 1992.
- [65] C. B. Saper, J. Lu, T. C. Chou, and J. Gooley, "The hypothalamic integrator for circadian rhythms," *Trends in Neurosciences*, vol. 28, no. 3, pp. 152–157, 2005.
- [66] K. Okamoto-Mizuno and K. Mizuno, "Effects of thermal environment on sleep and circadian rhythm," *Journal of Physiological Anthropology*, vol. 31, no. 1, pp. 1–9, 2012.
- [67] A. C. Verceles, L. Silhan, M. Terrin, G. Netzer, C. Shanholtz, and S. M. Scharf, "Circadian rhythm disruption in severe sepsis: The effect of ambient light on urinary 6-sulfatoxymelatonin secretion," *Intensive Care Medicine*, vol. 38, no. 5, pp. 804–810, 2012.
- [68] C. Blume, C. Garbazza, and M. Spitschan, "Effects of light on human circadian rhythms, sleep and mood," *Somnologie*, vol. 23, no. 3, pp. 147–156, 2019.
- [69] J. DuBose, R. G. Davis, G. Campiglia, A. Wilkerson, and C. Zimring, "Lighting the patient room of the future: Evaluating different lighting conditions from the patient perspective," *HERD: Health Environments Research & Design Journal*, p. 19 375 867 211 063 481, 2021.
- [70] A. Rashed, A. Ibrahim, A. Adel, B. Mourad, A. Hatem, M. Magdy, N. Elgaml, and A. Khattab, "Integrated iot medical platform for remote healthcare and assisted living," in *2017 Japan-Africa Conference on Electronics, Communications and Computers (JAC-ECC)*, IEEE, 2017, pp. 160–163.
- [71] A. Sato, M. Nakajima, and N. Kohtake, "Rapid ble beacon localization with range-only ekf-slam using beacon interval constraint," in *2019 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, IEEE, 2019, pp. 1–8.
- [72] A. J. Martín, I. M. Gordo, J. J. G. Domínguez, J. Torres-Sospedra, S. L. Plaza, and D. G. Gómez, "Affinity propagation clustering for older adults daily routine estimation," in *2021 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, IEEE, pp. 1–7.
- [73] D. Jain, K. Mack, A. Amrous, M. Wright, S. Goodman, L. Findlater, and J. E. Froehlich, "Homesound: An iterative field deployment of an in-home sound

awareness system for deaf or hard of hearing users,” in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–12.

- [74] M. Cantarini, A. Brocanelli, L. Gabrielli, and S. Squartini, “Acoustic features for deep learning-based models for emergency siren detection: An evaluation study,” in *2021 12th International Symposium on Image and Signal Processing and Analysis (ISPA)*, IEEE, 2021, pp. 47–53.
- [75] S. Wyatt, D. Elliott, A. Aravamudan, C. E. Otero, L. D. Otero, G. C. Anagnostopoulos, A. O. Smith, A. M. Peter, W. Jones, S. Leung, *et al.*, “Environmental sound classification with tiny transformers in noisy edge environments,” in *2021 IEEE 7th World Forum on Internet of Things (WF-IoT)*, IEEE, 2021, pp. 309–314.
- [76] N. A. Collop, W. M. Anderson, B. Boehlecke, D. Claman, R. Goldberg, D. J. Gottlieb, D. Hudgel, M. Sateia, and R. Schwab, “Clinical guidelines for the use of unattended portable monitors in the diagnosis of obstructive sleep apnea in adult patients,” *Journal of Clinical Sleep Medicine*, vol. 3, no. 7, pp. 737–747, 2007.
- [77] N. M. Punjabi, “The epidemiology of adult obstructive sleep apnea,” *Proceedings of the American Thoracic Society*, vol. 5, no. 2, pp. 136–143, 2008.
- [78] S. Pallesen, J. Grønli, K. Myhre, F. Moen, B. Bjorvatn, I. Hanssen, and H. S. A. Heglum, “A pilot study of impulse radio ultra wideband radar technology as a new tool for sleep assessment,” *Journal of Clinical Sleep Medicine*, vol. 14, no. 7, pp. 1249–1254, 2018.
- [79] T. Penzel, C. Schöbel, and I. Fietze, “New technology to assess sleep apnea: Wearables, smartphones, and accessories,” *F1000Research*, vol. 7, 2018.
- [80] T. D. Rosa, J. Zitser, and R. Capasso, “Consumer technology for sleep-disordered breathing: A review of the landscape,” *Current Otorhinolaryngology Reports*, pp. 1–9, 2019.
- [81] W. W. Flemons, N. J. Douglas, S. T. Kuna, D. O. Rodenstein, and J. Wheatley, “Access to diagnosis and treatment of patients with suspected sleep apnea,” *American Journal of Respiratory and Critical Care Medicine*, vol. 169, no. 6, pp. 668–672, 2004.
- [82] L. J. Epstein, D. Kristo, P. J. Strollo, N. Friedman, A. Malhotra, S. P. Patil, K. Ramar, R. Rogers, R. J. Schwab, E. M. Weaver, and M. D. Weinstein, “Clinical guideline for the evaluation, management and long-term care of obstructive sleep apnea in adults,” *Journal of Clinical Sleep Medicine*, vol. 5, no. 3, pp. 263–276, 2009.

- [83] A. Roebuck, V. Monasterio, E. Geder, M. Osipov, J. Behar, A. Malhotra, T. Penzel, and G. Clifford, "A review of signals used in sleep analysis," *Physiological measurement*, vol. 35, no. 1, R1, 2013.
- [84] G. Memis and M. Sert, "Multimodal classification of obstructive sleep apnea using feature level fusion," in *2017 IEEE 11th International Conference on Semantic Computing (ICSC)*, IEEE, 2017, pp. 85–88.
- [85] M. Deviaene, "Multimodal signal analysis for unobtrusive characterization of obstructive sleep apnea," Ph.D. dissertation, 2020.
- [86] J. Behar, A. Roebuck, J. S. Domingos, E. Geder, and G. D. Clifford, "A review of current sleep screening applications for smartphones," *Physiological measurement*, vol. 34, no. 7, R29, 2013.
- [87] C.-W. Wang, A. Hunter, N. Gravill, and S. Matusiewicz, "Unconstrained video monitoring of breathing behavior and application to diagnosis of sleep apnea," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 2, pp. 396–404, 2013.
- [88] G. D. Clifford and E. Geder, "Out of touch: From audio recordings to phone apps to mattress sensors, noncontact systems offer a less cumbersome way to monitor sleep.," *IEEE pulse*, vol. 5, no. 5, pp. 19–21, 2014.
- [89] E. Geder, "Video and audio analysis for the detection of obstructive sleep apnoea," Ph.D. dissertation, University of Oxford, 2017.
- [90] X. Liu, J. A. Hubbard, R. A. Fabes, and J. B. Adam, "Sleep disturbances and correlates of children with autism spectrum disorders," *Child psychiatry and human development*, vol. 37, no. 2, pp. 179–191, 2006.
- [91] S. Gunes, O. Ekin, A. Feyzioglu, N. Ekin, and M. Kalinli, "Sleep problems in children with autism spectrum disorder: Clinical correlates and the impact of attention deficit hyperactivity disorder," *Neuropsychiatric disease and treatment*, vol. 15, p. 763, 2019.
- [92] W. Chen, A. Sano, D. L. Martinez, S. Taylor, A. W. McHill, A. J. Phillips, L. Barger, E. B. Klerman, and R. W. Picard, "Multimodal ambulatory sleep detection," in *2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, IEEE, 2017, pp. 465–468.
- [93] A. S. Cakmak, G. Da Poian, A. Willats, A. Haffar, R. Abdalbaki, Y.-A. Ko, A. J. Shah, V. Vaccarino, D. L. Bliwise, C. Rozell, *et al.*, "An unbiased, efficient sleep-wake detection algorithm for a population with sleep disorders: Change point decoder," *Sleep*, vol. 43, no. 8, zsa011, 2020.

- [94] Q. Li, Q. Li, A. S. Cakmak, G. Da Poian, D. L. Bliwise, V. Vaccarino, A. J. Shah, and G. D. Clifford, "Transfer learning from ecg to ppg for improved sleep staging from wrist-worn wearables," *Physiological Measurement*, vol. 42, no. 4, p. 044 004, 2021.
- [95] E. Dafna, A. Tarasiuk, and Y. Zigel, "Sleep-wake evaluation from whole-night non-contact audio recordings of breathing sounds," *PloS one*, vol. 10, no. 2, e0117382, 2015.
- [96] A. Heinrich, X. Aubert, and G. de Haan, "Body movement analysis during sleep based on video motion estimation," in *2013 IEEE 15th International Conference on e-Health Networking, Applications and Services (Healthcom 2013)*, IEEE, 2013, pp. 539–543.
- [97] T. Rahman, A. T. Adams, R. V. Ravichandran, M. Zhang, S. N. Patel, J. A. Kientz, and T. Choudhury, "Dopplesleep: A contactless unobtrusive sleep sensing system using short-range doppler radar," in *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*, 2015, pp. 39–50.
- [98] M. Dixon, L. Schneider, J. Yu, J. Hsu, A. Pathak, D. Shin, R. S. Lee, M. R. Malhotra, K. Mixer, M. McConnell, *et al.*, "Sleep-wake detection with a contactless, bedside radar sleep sensing system," 2021.
- [99] J. Comstock, *Itamar gets FDA-clearance for disposable wearable that tests for sleep apnea*, [Online; accessed 18-Aug-2022].
- [100] *Withings*, [Online; accessed 18-Aug-2022].
- [101] *Beddit*, [Online; accessed 18-Aug-2022].
- [102] *UK Medical Device Company Acurable Receives US FDA Clearance for Its Home Sleep Apnea Testing Device AcuPebble SA100*, [Online; accessed 18-Aug-2022].
- [103] *RESONEA Obtains FDA Clearance for Smartphone-Based In-Home Screening for Obstructive Sleep Apnea*, [Online; accessed 18-Aug-2022].
- [104] *Sleep Apnea Detection App SleepCheckRx Receives FDA 510(k) Clearance*, [Online; accessed 18-Aug-2022].
- [105] M. L. McHugh, "Interrater reliability: The kappa statistic," *Biochemia medica*, vol. 22, no. 3, pp. 276–282, 2012.

- [106] Y. Hoshino, H. Watanabe, Y. Yashima, M. Kaneko, and H. Kumashiro, "An investigation on sleep disturbance of autistic children," *Psychiatry and Clinical Neurosciences*, vol. 38, no. 1, pp. 45–51, 1984.
- [107] C. R. Johnson, "Sleep problems in children with mental retardation and autism," *Child and adolescent psychiatric clinics of North America*, vol. 5, no. 3, pp. 673–684, 1996.
- [108] S. Tordjman, K. S. Davlantis, N. Georgieff, M.-M. Geoffray, M. Speranza, G. M. Anderson, J. Xavier, M. Botbol, C. Oriol, E. Bellissant, *et al.*, "Autism as a disorder of biological and behavioral rhythms: Toward new therapeutic perspectives," *Frontiers in Pediatrics*, vol. 3, p. 1, 2015.
- [109] S. Cohen, R. Conduit, S. W. Lockley, S. M. Rajaratnam, and K. M. Cornish, "The relationship between sleep and behavior in autism spectrum disorder (asd): A review," *Journal of neurodevelopmental disorders*, vol. 6, no. 1, pp. 1–10, 2014.
- [110] M. Chahrour and H. Y. Zoghbi, "The story of rett syndrome: From clinic to neurobiology.," *Neuron*, vol. 56, pp. 422–37, 3 Nov. 2007.
- [111] D. C. Tarquinio and A. K. Percy, *Rett syndrome: Clinical aspects*, 2016.
- [112] C. A. Chapleau, J. Lane, L. Pozzo-Miller, and A. K. Percy, "Evaluation of current pharmacological treatment options in the management of rett syndrome: From the present to future therapeutic alternatives," *Current clinical pharmacology*, vol. 8, pp. 358–369, 4 2013.
- [113] J. L. Neul, D. G. Glaze, A. K. Percy, T. Feyma, A. Beisang, T. Dinh, B. Suter, E. Anagnostou, M. Snape, J. Horrigan, *et al.*, "Improving treatment trial outcomes for rett syndrome: The development of rett-specific anchors for the clinical global impression scale," *Journal of child neurology*, vol. 30, pp. 1743–1748, 13 2015.
- [114] J. Busner and S. D. Targum, "The clinical global impressions scale: Applying a research tool in clinical practice.," *Psychiatry (Edgmont (Pa. : Township))*, vol. 4, pp. 28–37, 7 Jul. 2007.
- [115] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet," *Circulation*, vol. 101, 23 Jun. 2000.
- [116] J. Lee, S. Nemati, I. Silva, B. A. Edwards, J. P. Butler, and A. Malhotra, "Transfer entropy estimation and directional coupling change detection in

- biomedical time series," *Biomedical engineering online*, vol. 11, pp. 1–17, 1 2012.
- [117] S. P. Shashikumar, Q. Li, G. D. Clifford, and S. Nemati, "Multiscale network representation of physiological time series for early prediction of sepsis," *Physiological measurement*, vol. 38, p. 2235, 12 2017.
- [118] D. C. Tarquinio, *Treatment of mitochondrial dysfunction in rett syndrome with triheptanoin*, [Online; accessed 01-Apr-2021].
- [119] *Outcome measures and biomarkers development initiative*, [Online; accessed 01-Apr-2021].
- [120] C. Burns, *Mc10 "biostamp" connects to your body, shares data*, [Online; accessed 01-Apr-2021].
- [121] F. Azuaje, G. D. Clifford, and P. McSharry, *Advanced methods and tools for ECG data analysis*. 2006.
- [122] A. N. Vest, G. D. Poian, Q. Li, C. Liu, S. Nemati, A. J. Shah, and G. D. Clifford, "An open source benchmarked toolbox for cardiovascular waveform and interval analysis," *Physiological measurement*, vol. 39, p. 105 004, 10 2018.
- [123] Q. Li, R. G. Mark, and G. D. Clifford, "Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a kalman filter," *Physiological measurement*, vol. 29, no. 1, p. 15, 2007.
- [124] A. S. Cakmak and G. D. Clifford, *Open source actigraphy toolbox*, [Online; accessed 01-Apr-2021], 2020.
- [125] M. Borazio, E. Berlin, N. Kücükıldiz, P. Scholl, and K. V. Laerhoven, "Towards benchmarked sleep detection with wrist-worn sensing units," 2014, pp. 125–134.
- [126] G. D. Clifford, F. Azuaje, P. McSharry, *et al.*, *Advanced methods and tools for ECG data analysis*. Artech house Boston, 2006, vol. 10.
- [127] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of biological signals," *Physical review E*, vol. 71, no. 2, p. 021 906, 2005.
- [128] S. Nemati, B. A. Edwards, J. Lee, B. Pittman-Polletta, J. P. Butler, and A. Malhotra, "Respiration and heart rate complexity: Effects of age and gender assessed by band-limited transfer entropy," *Respiratory physiology & neurobiology*, vol. 189, pp. 27–33, 1 2013.

- [129] D. M. Katz, A. Bird, M. Coenraads, S. J. Gray, D. U. Menon, B. D. Philpot, and D. C. Tarquinio, "Rett syndrome: Crossing the threshold to clinical translation," *Trends in neurosciences*, vol. 39, pp. 100–113, 2 2016.
- [130] A. Bauer, J. W. Kantelhardt, P. Barthel, R. Schneider, T. Mäkikallio, K. Ulm, K. Hnatkova, A. Schömig, H. Huikuri, A. Bunde, M. Malik, and G. Schmidt, "Deceleration capacity of heart rate as a predictor of mortality after myocardial infarction: Cohort study," *The Lancet*, vol. 367, pp. 1674–1681, 9523 May 2006.
- [131] A. J. Ma, N. Rawat, A. Reiter, C. Shrock, A. Zhan, A. Stone, A. Rabiee, S. Griffin, D. M. Needham, and S. Saria, "Measuring patient mobility in the ICU using a novel noninvasive sensor," *Critical care medicine*, vol. 45, no. 4, p. 630, 2017.
- [132] A. Davoudi, K. R. Malhotra, B. Shickel, S. Siegel, S. Williams, M. Ruppert, E. Bihorac, T. Ozrazgat-Baslanti, P. J. Tighe, A. Bihorac, *et al.*, "Intelligent ICU for autonomous patient monitoring using pervasive sensing and deep learning," *Scientific reports*, vol. 9, no. 1, pp. 1–13, 2019.
- [133] F. Schmid, M. S. Goepfert, and D. A. Reuter, "Patient monitoring alarms in the ICU and in the operating room," *Annual Update in Intensive Care and Emergency Medicine 2013*, pp. 359–371, 2013.
- [134] E. Ryherd, J. Ackerman, C. Zimring, S. Okcu, and K. Persson Waye, "Noise pollution in hospitals: Impacts on staff," *Journal of clinical outcomes management: JCOM*, vol. 19, pp. 1–10, Nov. 2012.
- [135] D. B. Choiniere, "The effects of hospital noise," *Nursing administration quarterly*, vol. 34, no. 4, pp. 327–333, 2010.
- [136] S. J. Jaiswal, S. Garcia, and R. L. Owens, "Sound and light levels are similarly disruptive in ICU and non-ICU wards.," *Journal of hospital medicine*, vol. 12, no. 10, pp. 798–804, 2017.
- [137] J. Wise, "Higher nurse to patient ratio is linked to reduced risk of inpatient death," *BMJ*, vol. 352, 2016.
- [138] J. Frisby, V. Smith, S. Traub, and V. L. Patel, "Contextual computing: A Bluetooth based approach for tracking healthcare providers in the emergency room," *Journal of biomedical informatics*, vol. 65, pp. 97–104, 2017.
- [139] A. Vankipuram and V. L. Patel, "Automated location tracking in clinical environments: A review of systems and impact on workflow analysis," in *Cognitive Informatics*, Springer, 2019, pp. 235–253.

- [140] R. D. White, "Recommended standards for the newborn ICU," *Journal of Perinatology*, vol. 27, no. 2, S4–S19, 2007.
- [141] J. Wilde-Frenz and H. Schulz, "Rate and distribution of body movements during sleep in humans," *Perceptual and motor skills*, vol. 56, no. 1, pp. 275–283, 1983.
- [142] P. B. Suresha, A. S. Cakmak, G. Da Poian, A. J. Shah, V. Vaccarino, D. Bliwise, and G. D. Clifford, "Obstructive sleep apnea classification in a mixed-disorder elderly male population using a low-cost off-body movement sensor," in *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, IEEE, 2019, pp. 1–4.
- [143] P.-C. Liang and P. Krause, "Real-time indoor patient movement pattern telemonitoring with one-meter precision," in *2014 4th International Conference on Wireless Mobile Communication and Healthcare-Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH)*, IEEE, 2014, pp. 141–144.
- [144] C. R. Baker, K. Armijo, S. Belka, M. Benhabib, V. Bhargava, N. Burkhart, A. Der Minassians, G. Dervisoglu, L. Gutnik, M. B. Haick, *et al.*, "Wireless sensor networks for home health care," in *21st International Conference on Advanced Information Networking and Applications Workshops (AINAW'07)*, IEEE, vol. 2, 2007, pp. 832–837.
- [145] S. H. Kim, J. M. Jeong, M. T. Hwang, and C. S. Kang, "Development of an iot-based atmospheric environment monitoring system," in *2017 International Conference on Information and Communication Technology Convergence (ICTC)*, IEEE, 2017, pp. 861–863.
- [146] F. Deng, J. Dong, X. Wang, Y. Fang, Y. Liu, Z. Yu, J. Liu, and F. Chen, "Design and implementation of a noncontact sleep monitoring system using infrared cameras and motion sensor," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 7, pp. 1555–1563, 2018.
- [147] C. Hegde, Z. Jiang, P. B. Suresha, J. Zelko, S. Seyedi, M. A. Smith, D. W. Wright, R. Kamaleswaran, M. A. Reyna, and G. D. Clifford, "Autotriage - an open source edge computing raspberry pi-based clinical screening system," *medRxiv*, 2020.
- [148] M. D. Rienzo and R. Mukkamala, "Wearable and nearable biosensors and systems for healthcare," *Wearable and Nearable Biosensors and Systems for Healthcare*, 2021.

- [149] R. Romdhane, E. Mulin, A. Derreumeaux, N. Zouba, J. Piano, L. Lee, I. Leroi, P. Mallea, R. David, M. Thonnat, *et al.*, “Automatic video monitoring system for assessment of alzheimer’s disease symptoms,” *The Journal of Nutrition, Health & Aging*, vol. 16, no. 3, pp. 213–218, 2012.
- [150] S. Suzuki, T. Matsui, M. Kagawa, T. Asao, and K. Kotani, “An approach to a non-contact vital sign monitoring using dual-frequency microwave radars for elderly care,” *Journal of Biomedical Science and Engineering*, vol. 6, no. 7, 2013.
- [151] T. Matsui, Y. Yoshida, M. Kagawa, M. Kubota, and A. Kurita, “Development of a practicable non-contact bedside autonomic activation monitoring system using microwave radars and its clinical application in elderly people,” *Journal of Clinical Monitoring and Computing*, vol. 27, no. 3, pp. 351–356, 2013.
- [152] N. Tsumura, C. Kanai, K. Masui, R. Mitsuhashi, T. Nagasawa, *et al.*, “Automatic classification of adult males with and without autism spectrum disorder by non-contact measurement of autonomic nervous system activation,” *Frontiers in Psychiatry*, vol. 12, p. 573, 2021.
- [153] A. Shajahan, C. H. Culp, and B. Williamson, “Effects of indoor environmental parameters related to building heating, ventilation, and air conditioning systems on patients’ medical outcomes: A review of scientific research on hospital buildings,” *Indoor Air*, vol. 29, no. 2, pp. 161–176, 2019.
- [154] M. A. Melhado, J. Hensen, and M. Loomans, “Literature review of staff thermal comfort and patient thermal risks in operating rooms,” in *8th International healthy buildings conference*, 2006, pp. 11–14.
- [155] J. W. Tang, “The effect of environmental parameters on the survival of airborne infectious agents,” *Journal of the Royal Society Interface*, vol. 6, no. suppl.-6, S737–S746, 2009.
- [156] P. B. Suresha and G. D. Clifford, *OS-Edge-Compute-Data-Capture-RPi*, <https://github.com/cliffordlab/OS.Edge.Compute.Data.Capture.RPi>, 2021.
- [157] *Pi NoIR Camera V2*, <https://www.raspberrypi.org/products/pi-noir-camera-v2/>, [Online; accessed 24-Feb-2021].
- [158] *Python-sounddevice*, <https://python-sounddevice.readthedocs.io/>, [Online; accessed 24-Feb-2021].
- [159] *Github: Python-sounddevice*, <https://github.com/spatialaudio/python-sounddevice/>, [Online; accessed 24-Feb-2021].

- [160] *Librosa*, <https://librosa.org/doc/latest/index.html>, [Online; accessed 24-Feb-2021].
- [161] *mse.c*, <https://www.physionet.org/physiotools/mse/mse.c>, [Online; accessed 24-Feb-2021].
- [162] *Bluepy*, <https://ianharvey.github.io/bluepy-doc/>, [Online; accessed 24-Feb-2021].
- [163] C. Hegde, R. Sameni, A. B. Rad, and G. D. Clifford, "Modeling social distancing and quantifying epidemic disease exposure in a built environment," *IEEE Journal of Selected Topics in Signal Processing*, pp. 1–1, 2022.
- [164] M. Tsai, A. M. Mori, C. W. Forsberg, N. Waiss, J. L. Sporleder, N. L. Smith, and J. Goldberg, "The Vietnam era twin registry: A quarter century of progress," *Twin Research and Human Genetics*, vol. 16, no. 1, pp. 429–436, 2013.
- [165] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of complex physiologic time series," *Physical Review Letters*, vol. 89, no. 6, p. 068 102, 2002.
- [166] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, no. Mar, pp. 1157–1182, 2003.
- [167] F. E. Block Jr, J. D. Rouse, M. Hakala, and C. L. Thompson, "A proposed new set of alarm sounds which satisfy standards and rationale to encode source information," *Journal of clinical monitoring and computing*, vol. 16, no. 7, p. 541, 2000.
- [168] *ISO/IEC 60601-1-8, Patterson and other alarms in medical equipment Sample Alarm Sounds - Sirens, buzzers and other sounds*, <http://www.anaesthesia.med.usyd.edu.au/resources/alarms/>, [Online; accessed 02-Jun-2020].
- [169] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.
- [170] *XGBoost*, <https://xgboost.readthedocs.io/en/latest/index.html>, [Online; accessed 24-Feb-2021].
- [171] K. Takahashi, K. Yamamoto, A. Kuchiba, and T. Koyama, "Confidence interval for micro-averaged F1 and macro-averaged F1 scores," *Applied Intelligence*, pp. 1–12, 2021.

- [172] T. Afouras, J. S. Chung, and A. Zisserman, "The conversation: Deep audio-visual speech enhancement," in *INTERSPEECH*, 2018.
- [173] C. Wallace, J. Robins, L. Alvord, and J. Walker, "The effect of earplugs on sleep measures during exposure to simulated intensive care unit noise," *American Journal of Critical Care*, vol. 8, no. 4, pp. 210–219, Jul. 1999.
- [174] G. L. Weinhouse and P. L. Watson, "Sedation and sleep disturbances in the ICU," *Critical Care Clinics*, vol. 25, no. 3, pp. 539–549, 2009.
- [175] A. Mitchell, K. Gudeczauskas, A. Therrien, and A. Zauher, "Bedside reporting is a key to communication," *Journal of Healthcare Communications*, vol. 3, no. 1, p. 13, 2018.
- [176] N. Oakley, "Validation with polysomnography of the sleepwatch sleep/wake scoring algorithm used by the actiwatch activity monitoring system," *Bend: Mini Mitter, Cambridge Neurotechnology*, 1997.
- [177] M. Borazio, E. Berlin, N. Kücükıldiz, P. Scholl, and K. Van Laerhoven, "Towards benchmarked sleep detection with wrist-worn sensing units," in *2014 IEEE International Conference on Healthcare Informatics*, IEEE, 2014, pp. 125–134.
- [178] A. Kosmadopoulos, C. Sargent, D. Darwent, X. Zhou, and G. D. Roach, "Alternatives to polysomnography (psg): A validation of wrist actigraphy and a partial-psg system," *Behavior research methods*, vol. 46, no. 4, pp. 1032–1041, 2014.
- [179] A. Khaustov, S. Nemati, and G. Clifford, "An open-source standard t-wave alternans detector for benchmarking," in *2008 Computers in Cardiology*, IEEE, 2008, pp. 509–512.
- [180] A. N. Vest, G. Da Poian, Q. Li, C. Liu, S. Nemati, A. J. Shah, and G. D. Clifford, "An open source benchmarked toolbox for cardiovascular waveform and interval analysis," *Physiological measurement*, vol. 39, no. 10, p. 105 004, 2018.
- [181] Q. Li, Q. Li, C. Liu, S. P. Shashikumar, S. Nemati, and G. D. Clifford, "Deep learning in the cross-time frequency domain for sleep staging from a single-lead electrocardiogram," *Physiological measurement*, vol. 39, no. 12, p. 124 005, 2018.
- [182] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "Catboost: Unbiased boosting with categorical features," *Advances in neural information processing systems*, vol. 31, 2018.

- [183] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and psychological measurement*, vol. 20, no. 1, pp. 37–46, 1960.
- [184] S. F. Quan, B. V. Howard, C. Iber, J. P. Kiley, F. J. Nieto, G. T. O'Connor, D. M. Rapoport, S. Redline, J. Robbins, J. M. Samet, *et al.*, "The sleep heart health study: Design, rationale, and methods," *Sleep*, vol. 20, no. 12, pp. 1077–1085, 1997.
- [185] L. Prechelt, "Early stopping-but when?" In *Neural Networks: Tricks of the trade*, Springer, 1998, pp. 55–69.
- [186] X. Long, R. Otte, E. v. d. Sanden, J. Werth, and T. Tan, "Video-based actigraphy for monitoring wake and sleep in healthy infants: A laboratory study," *sensors*, vol. 19, no. 5, p. 1075, 2019.
- [187] M. Awais, X. Long, B. Yin, S. F. Abbasi, S. Akbarzadeh, C. Lu, X. Wang, L. Wang, J. Zhang, J. Dudink, *et al.*, "A hybrid dcnn-svm model for classifying neonatal sleep and wake states based on facial expressions in video," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 5, pp. 1441–1449, 2021.
- [188] X. Long, J. Espina, R. A. Otte, W. Wang, R. M. Aarts, and P. Andriessen, "Video-based actigraphy is an effective contact-free method of assessing sleep in preterm infants," *Acta Paediatrica (Oslo, Norway: 1992)*, vol. 110, no. 6, p. 1815, 2021.