

A Computational Framework for Modeling Cyclic Human Behaviors from Multi-Modal Sensor Data

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

in partial fulfillment
of the requirements for the degree

Doctor of Philosophy

by

Runze Yan

May 2023

APPROVAL SHEET

This
Dissertation
is submitted in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

Author: Runze Yan

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

Advisor: Afsaneh Doryab

Advisor:

Committee Member: Laura Barnes

Committee Member: Afsaneh Doryab

Committee Member: Philip I. Chow

Committee Member: Donald E. Brown

Committee Member: Gregory J. Gerling

Committee Member:

Accepted for the School of Engineering and Applied Science:



Jennifer L. West, School of Engineering and Applied Science

May 2023

A Computational Framework for Modeling Cyclic Human Behaviors from Multi-Modal Sensor Data

Runze Yan

(ABSTRACT)

Cyclic human behaviors, such as circadian rhythms, have significant implications for health outcomes. These rhythms have primarily been studied in laboratory and controlled settings with small and sparse amount of data collected in short periods. With the advancements in mobile and wearable devices, it is possible to collect longitudinal, continuous, and fine-grained biobehavioral data from individuals in the wild. While this data enables rigorous modeling of cycling human behavior, existing time series modeling techniques are insufficient for this task as they assume that the input time series is strictly stationary, are unable to independently learn non-dominant cycles, require the number and types of cyclic patterns, and cannot process massive sensor data. This dissertation presents a novel computational framework for modeling cyclic human behaviors from multi-modal mobile sensing data. The developed framework is designed to be adaptable in processing data with diverse time granularity and is able to automatically detect and model cyclic human behaviors from multi-modal mobile sensing data. Moreover, it can model similarities and differences in rhythms across individuals and time, predict various outcomes related to well-being and health, identify activity sub-patterns for a single person or across a population, and facilitate an interpretable description of human behavior. Importantly, this is the first framework that addresses the variability of cyclic human behaviors using non-stationary time series. To evaluate the effectiveness of the proposed framework,

both open-source and self-collected mobile sensing datasets are used. Through this evaluation, the framework's ability to process multi-modal mobile sensing data and accurately model cyclic human behaviors is demonstrated. Overall, this work has the potential to significantly advance the understanding of cyclic human behaviors and their relationship to health outcomes in real-world settings.

Acknowledgments

I would like to express my deepest gratitude to my advisor, Dr. Afsaneh Droayb, for her unwavering guidance and support throughout my PhD journey. Her invaluable insights, encouragement, and feedback have been instrumental in shaping my research and helping me achieve my goals. I would also like to extend my heartfelt thanks to my committee members for their valuable feedback and suggestions that have greatly improved my work. Their expertise and insights have been invaluable in shaping the direction of my research. I am also grateful to the members of the Human-AI Technology (HAI) Lab for their support and assistance during the preparation of my thesis and presentation. Their dedication, expertise, and encouragement have been instrumental in helping me achieve my goals. I would also like to thank my family and friends for their unwavering support throughout my PhD journey. Their encouragement, love, and support have been a source of strength and inspiration throughout this challenging journey. Finally, I would like to express my gratitude to the Systems and Information Engineering Department and the Link Lab at the University of Virginia for providing me with the opportunity to pursue my PhD and for their unwavering support throughout my academic journey.

Contents

1	Introduction	1
2	Literature Review	5
2.1	Cyclic Human Behaviors	5
2.2	Time Series Algorithms for Human Behavior Analysis	7
2.2.1	Time Series Models for Human Behavior Modeling	8
2.2.2	Processes Specifically for Modeling Cyclic Human Behavior	12
2.2.3	Modeling Human Behavior from Mobile Sensor Data	16
2.3	Image Representation of Sensor Data	18
2.3.1	Encode Time Series into Image	18
2.3.2	Transfer Learning for Image Processing	19
3	CoRhythmo Framework for Modeling Cyclic Human Behaviors	21
3.1	Automated Pipeline of Processing Multi-modal Sensor Data for Modeling Human Behaviors	23
3.2	Approaches for Modeling Rhythms from Stationary Time Series	26
3.2.1	Periodicity Detection from Stationary Time series	27
3.2.2	Rhythm Characteristic Extraction from Stationary Time Series	30

3.3	A Novel Wavelet Transfer Learning (WTL) Method for Modeling Rhythms from Non-stationary Time Series	32
3.3.1	Multi-modal Time Series Wavelet Transform	33
3.3.2	Behavior Modeling and Interpretation via Fused Wavelet Spectrum Transfer Learning	38
3.3.3	Granular Behavior Inspection with Power Spectrum Segmentation	40
3.4	Multi-dimensional Correlation Identification in Rhythm Analysis	41
3.5	Prediction of Outcomes	46
3.5.1	Feature Selection	47
3.5.2	Model Building and Validation	48
3.6	Rhythm Variability Modeling	49
3.6.1	Changing Point Detection in Cycles	50
3.6.2	A Noval Stability Measurement for Human Rhythms from Sensor Streams	52
4	Applications of the Framework	59
4.1	Predicting Human Outcomes Based on Rhythm Modeling	59
4.1.1	Case1: Predicting Mental Health	59
4.1.2	Case2: Predicting Readiness	78
4.2	Identifying Relationship between Rhythms and Productivity	83

4.2.1	Correlation Aggregation of Rhythm Parameters	85
4.2.2	Correlation Aggregation of Sensor Features	89
4.3	Measuring Variability in Biobehavioral Rhythms	91
4.3.1	Identifying Rhythm Disruption in Physiological Cycles	91
4.3.2	Quantifying Rhythm Variability of Readiness and Well-being Outcomes	95
4.4	Mining Sub-patterns for Human Activity Recognition with WTL . . .	103
4.4.1	Activity Clustering	104
4.4.2	Behavioral Sub-patterns	107
4.4.3	Granular Behavior Inspection	108
5	Discussion	113
5.1	Limitations	115
5.2	Future Work	116
6	Conclusions	118
	Appendices	145
	Appendix A Evaluation of the State-of-the-Art Periodicity Detection	
	Methods	146
A.1	Datasets	146
A.2	Periodicity Detection	148

Chapter 1

Introduction

Numerous research has examined periodic phenomena in living organisms and confirmed that exploring the cyclic human behaviors is an effective way to diagnose and treat many illnesses such as cancer, cardiovascular disease, and mental health problems [4, 30, 81]. For example, patients with depression, bipolar disorder, and schizophrenia usually exhibit irregular changes in circadian rhythm, and adjusting the circadian rhythm is an efficient auxiliary method for treating these conditions [106, 64]. Disruption in biological rhythms is also caused by changing lifestyles and environmental conditions such as travel across time zones and shift work [59]. For example, sleep disorders such as delayed sleep stage disorder (DSPD) and advanced sleep stage disorder (ASPD) are associated with circadian rhythms too late and too early [41, 57]. Jet lag is another common circadian rhythm disorder [59, 14]. The technology of using sensors of a mobile device (i.e. smartphone and wearable devices) to acquire data is named mobile sensing; such devices provide the capability of continuous and unobtrusive tracking of biobehavioral signals of individuals in their daily life and outside of controlled lab settings. Particularly, the current mobile devices have been able to record physiological and physical activity data effectively. For example, two studies with healthy young adults have used activity data from Fitbit devices to quantify rest-activity rhythms and found that rhythm measurement compared well relative to research-grade actigraphy [111, 145].

There have been numerous studies demonstrating the reliability of mobile sensors in capturing circadian patterns, including sleep, physical and health conditions. For example, studies using actigraphy devices have shown differences in circadian rhythms among patients with bipolar disorder, ADHD, and schizophrenia [34]. Abdullah et al. used patterns of phone usage to identify chronotypes of students (early birds or night owls) [98]. Murnane et al. aggregated mobile application usage features by body clock time and analyzed the correlation between circadian rhythms in application usage and alertness level [125]. Moreover, Doryab et al. demonstrated modeling of rhythms using data from Fitbit devices in cancer patients and showed that disruption in circadian rhythms predicts readmission in cancer patients undergoing treatment [153].

Although previous studies have investigated human rhythms, they have focused mainly on the 24-hour circadian rhythm and daily routines, overlooking the full range of natural rhythms in human behaviors. Mobile sensing technologies enable the collection of data from various sources, which has the potential to capture human behaviors more comprehensively. However, this also presents new challenges for cyclic time series modeling algorithms. The longitudinal time series data collected from personal devices is often noisy, incomplete, and voluminous, requiring careful processing to extract useful knowledge. Additionally, each data source captures different aspects of human behaviors, making it essential to incorporate and explore each signal to identify biological and behavioral indicators that reveal microscopic or macroscopic levels of cyclic behavior. Furthermore, multiple cyclic patterns may occur during a time interval, and the number of those patterns is unknown. Current cyclic time series modeling methods assume the time series are stable and ignore the variability of human behaviors, which can result in inaccurate results. Finally, generating an exhaustive number of rhythm models from different sensor sources in different time

granularities makes it challenging to interpret the models manually. To address these challenges, I propose a novel framework for modeling cyclic human behaviors from multi-modal mobile sensing data offers several significant contributions to the field of human behavior understanding, machine learning, and ubiquitous computing. Some key contributions are outlined below:

- Identification of cyclic patterns: The ability to identify and model cyclic patterns in human behavior is a significant advancement in the field. The framework can either automatically estimate the periods and shapes of the detected rhythms from stationary time series, or attribute semantic labels to behavioral patterns in an unsupervised manner from non-stationary time series. These assigned labels facilitate a more intuitive comprehension of a behavior’s characteristics while revealing sub-patterns within distinct human activity types, further enhancing the understanding of cyclic human behaviors.
- Enhanced behavior modeling: By leveraging multi-modal data from mobile sensing devices, the framework provides a more comprehensive and accurate representation of human behaviors. Additionally, the framework can flexibly process massive multi-modal sensor data in different time granularities for further human behavior modeling. This enables the identification of the multi-dimensional relationship between rhythms and human outcomes, and the detection of disruptions in these rhythms. Moreover, it allows for the quantification of the variability of rhythms between user groups (e.g., well-being or health groups) or across time. These insights are essential for enhancing overall human behavior understanding.
- Improved machine learning models: By accounting for cyclic patterns in human

behavior, the framework can lead to the development of more accurate and efficient machine learning models. These models can be used to predict outcomes related to health and well-being.

- Contributions to interdisciplinary research: The proposed framework demonstrates significant applicability in various interdisciplinary research contexts. Examples include predicting mental health status and readiness levels, investigating the connection between human rhythms and productivity, and measuring behavioral rhythm variability in relation to well-being factors such as cognitive, emotional, and physical energy levels. These applications yield new insights that would not be possible without the understanding of cyclic human behaviors derived from multi-modal sensing data, thereby showcasing the framework's academic value and potential impact.

In the following sections, I first offers a through examination of relevant research, and existing methodologies related to cyclic human behaviors and the analysis of multi-modal sensor data. Then, I detail the development of the novel computational framework, outlining the design, implementation, and validation of the proposed approach. In the results section, the findings obtained from applying the framework to both simulated and real-world data are presented and evaluated, highlighting its effectiveness in modeling biobehavioral rhythms. In the end, I discussed the broad significance and the timeline of the proposed work.

Chapter 2

Literature Review

Understanding and modeling cyclic human behaviors have become increasingly important as technology advances and integrates more deeply into the daily lives. Cyclic human behaviors, such as daily routines and seasonal activities, provide valuable insights into various aspects of human life, including health, productivity, and well-being. By accurately modeling these behaviors, researchers and practitioners can develop more effective, personalized, and context-aware systems, such as smart homes, healthcare monitoring solutions, and recommender systems. In this section, I will review the existing literature on cyclic human behaviors, , as well as time series algorithms for modeling human behavior. By doing so, I aim to provide a thorough understanding of the current state of the art, identify gaps and challenges in the field, and demonstrate how the proposed computational framework addresses these issues to advance the understanding and modeling of cyclic human behaviors from multi-modal sensor data. The review will still focus on image representation of sensor data, and the application of transfer learning for image processing in the context of cyclic human behavior modeling.

2.1 Cyclic Human Behaviors

The investigation of cyclic phenomena in living organisms uncovers the presence of recurring events and behaviors that follow specific cycles and can be represented

by periodic functions [17, 47]. Each periodic function is characterized by its mean level, oscillation magnitude, and optimal oscillation duration. Biological rhythms, such as activity and rest patterns or circadian rhythms, have been widely explored in Chronobiology and medicine [151, 80, 63], predominantly in controlled environments. One benefit of examining these cyclic behaviors is the improved understanding of how human physiology and behavior are influenced by internal and external factors [60]. This can help identify correlations and underlying mechanisms between human behavior and various health conditions [40], ultimately contributing to better diagnostic tools [178], treatments [51], and preventive measures. For example, studies in [151, 80, 63] demonstrate the association between long-term disruption in biological rhythms and health outcomes such as cancer, diabetes, and depression. Other studies have shown the impact of shift work on the quality of life in shift workers such as nurses and doctors [38, 136]. These studies, however, have often been limited to controlled settings to observe certain behaviors and effects. However, these studies have often been restricted to controlled settings to observe specific behaviors and effects. Technological advancements in activity trackers have facilitated the study of these phenomena beyond laboratory settings, exhibiting the reliability of such devices in detecting circadian disruptions related to sleep and physical and mental health conditions [195, 188, 194]. For example, studies utilizing research-grade actigraphy devices have identified differences in circadian rhythms among patients with bipolar disorder, ADHD, and schizophrenia [34]. Other research has employed the same data type to examine circadian disruption in cancer patients undergoing chemotherapy [34]. Commercial devices, such as Fitbits, are now capable of reasonably accurately deducing sleep duration and quality. Two concise studies involving healthy young adults utilized activity data from Fitbit devices to quantify rest-activity rhythms, discovering

that rhythm measurements correlated well with research-grade actigraphy [120, 138]. Investigations in [130] and [124] have also examined the ability of personal tracking devices to measure sleep in comparison to gold standard methods like polysomnography. In summary, analyzing cyclic human behaviors using mobile sensing technologies not only expands our understanding of biological rhythms but also fosters innovative approaches in monitoring, diagnosing, and treating various health conditions.

2.2 Time Series Algorithms for Human Behavior Analysis

In this section, the goal is to present a thorough summary of time series analysis techniques tailored for human behavior modeling. Human behavior modeling entails the representation of behavioral patterns derived from time series data [29]. This process uncovers hidden insights into human behaviors from raw or modified time series and has emerged as a vital research field across multiple disciplines, including psychology [109], social sciences [158], and computer science [160]. This review will first examine cutting-edge general time series models and assess their suitability for modeling human behaviors. Subsequently, an overview of time series methods specifically designed for cyclic time series will be provided, along with a discussion of their distinct advantages for modeling human behaviors compared to general time series models. Lastly, the current challenges in human behavior analysis using mobile sensing will be summarized.

2.2.1 Time Series Models for Human Behavior Modeling

The field of human behavior modeling has attracted considerable interest owing to its potential impact across multiple domains. A key method for modeling human behavior involves the examination of time series data, which can be categorized as either stationary or non-stationary [165, 168]. Analyzing human behavior from these two types of time series data necessitates distinct approaches and considerations, given the unique properties of each data category [162]. In the subsequent paragraphs, a discussion of time series for both stationary and non-stationary data, along with their successful implementation in human behavior modeling, will be presented.

Stationary time series data, by definition, exhibit constant statistical properties over time, such as mean and variance [54]. This allows for simpler modeling techniques, such as autoregressive integrated moving average (ARIMA) [25], exponential smoothing state space models (ETS) [42], and seasonal decomposition of time series (STL) [13], which rely on linear relationships and decomposable components. Autoregressive (AR) and moving average (MA) models have been commonly used to model human behavior patterns. Box and Jenkins [25] introduced the ARIMA model, which combines AR, MA, and integrated components. ARIMA models have been utilized in numerous studies, such as in the work of Matarazzo et al. [10], where they analyzed conversational patterns between therapists and clients. STL is an important technique in time series analysis, which is useful in capturing seasonal patterns in human behavior. Cleveland et al. [13] introduced the STL method, which has since been employed in various studies, including analyses of mood fluctuations, consumer behavior, and social media activity patterns [109, 89, 87, 118]. ETS is a class of time series forecasting methods that are based on the state space model framework [42]. ETS models are particularly useful when the time series exhibits trend, seasonality, and/or

irregularity [65]. ETS models have been applied to model various human behaviors, including forecasting web traffic (Hyndman et al. [65]) and modeling disease progression (De Livera et al. [78]). However, many real-world human behavior datasets exhibit non-stationary characteristics, such as time-varying mean or variance, which may limit the applicability of stationary time series models.

A non-stationary time series exhibits variations in its statistical properties [54]. One common approach to deal with non-stationary time series is differencing [6, 45], which is employed in the ARIMA model to transform the original data into a stationary form by calculating the differences between consecutive observations. This process can be applied iteratively until the resulting time series becomes stationary, allowing classical models like ARIMA to be effectively used for non-stationary human behavior data [149]. The decomposition of time series into its components, such as trend, seasonality, and residuals, can also help address non-stationarity [6]. STL, as mentioned earlier, is one such method used to extract and model seasonal patterns in non-stationary human behavior data [141].

State space transition models, such as Markov chains, are one class of models that have been used for modeling non-stationary time series in human behavior research [62]. State space transition models consist of a system equation, which describes the underlying state dynamics, and an observation equation, which links the state to the observable time series [37]. Rabiner et al. provided a comprehensive introduction to state space models, highlighting their flexibility in handling various types of non-stationary time series data [12]. Valstar et al. [55] employed hidden Markov models to analyze facial expressions.

More recent deep learning techniques, such as the recurrent neural network (RNN) and Transformer models, can handle non-stationary data more effectively [182]. These

models can adapt to the changing patterns in the data, making them suitable for modeling complex and dynamic human behavior patterns without explicit stationarization or differencing procedures [134, 156]. Recurrent neural networks (RNNs) represent a class of neural networks specifically designed for the analysis of sequential data, such as time series [186]. The architectural foundation of RNNs incorporates hidden states, which enable these networks to account for prior information effectively [180]. As a result, RNNs are particularly adept at modeling underlying patterns and dependencies within time series data, making them a fitting choice for tasks within this domain. Choi et al. [94] developed a RNN model specifically for predicting user behavior in software applications. The long short-term memory (LSTM) network, a type of RNN, have been applied by Tian et al. [185] for predicting user behavior in mobile applications. Transformer-based models have gained significant attention in recent years due to their ability to effectively model long-range dependencies in time series data [193]. Shafiullah et al. designed a new Transformer architecture to model unlabeled driving behavioral patterns [191]. Sun et al. developed a multi-modal Transformer-based framework to capture the long-range dependency of human behaviors for depression detection [192]. In summary, non-stationary models can better handle the changing statistical properties in time series data and model complex, non-linear relationships. Many human behaviors naturally exhibit non-stationary properties, making non-stationary time series models more appropriate for capturing and modeling these dynamics. However, non-stationary models can be more complex, requiring more parameters and a deeper understanding of the underlying algorithms. This may increase computational requirements and make interpretation more challenging. Due to the greater flexibility of non-stationary models, there may be an increased risk of overfitting, leading to poor generalization performance.

Furthermore, time series analysis methods have been combined with other approaches, such as feature extraction and dimensionality reduction, to better capture and model human behavior [188, 194]. Picard et al. integrated time series analysis with feature extraction techniques to assess emotions from physiological signals [39]. In another study, Wang et al. employed time series analysis and dimensionality reduction techniques to predict human actions from video data [88]. The parameters derived from time series models can be subsequently utilized as features within a prediction system. This approach effectively incorporates time series characteristics into the prediction process, enhancing its performance and accuracy. For example, authors in [181] used the LSTM model to process the time series data from ECG signals and learn the temporal relationship between ECG features and blood pressure values, which are further used for blood pressure prediction.

While the aforementioned time series methods offer a degree of utility in analyzing human behaviors, their applicability to investigating cyclic human behaviors may be limited. For instance, many conventional models, such as hidden Markov models and neural networks, primarily predict future values within time series data without effectively extracting cyclic patterns (e.g., period, amplitude, and phase). Consequently, these methods do not provide sufficient interpretability for cyclic human behaviors. Though the stationary ARIMA model can identify cyclic patterns, it fails to determine the periods for all underlying cycles. In addition, these existing time series methods may face difficulties when handling raw, multi-modal mobile sensor data. In particular, these multi-modal sensor datasets often exhibit large volumes, longitudinal structures, high dimensionality, noise, and missing values [187]. To address these issues, the developed framework in my work integrates various pre-processing components tailored for time series analysis, which include automated processing (e.g.,

data cleaning, missing value imputation, and temporal window segmentation) and modeling. The framework is generalizable and can incorporate any stationary and non-stationary time series method.

2.2.2 Processes Specifically for Modeling Cyclic Human Behavior

In this section, I will describe algorithms designed to detect cyclic patterns within time series data. This procedure, known as rhythm modeling, allows for improved comprehension and analysis of recurring human behaviors. Consistent rhythms are commonly observed in stationary time series, wherein the extraction of steady rhythmic characteristics can be achieved through periodicity detection and shape estimation of cyclic time series. In contrast, non-stationary time series exhibit fluctuating frequencies over time. The Wavelet transform serves as a useful tool to identify these variations in periods.

Periodicity Detection

Among the periodicity detection methods, Fast Fourier Transform (FFT) is the most commonly used approach. FFT converts a time function into a frequency function. The dominant frequency in the frequency function is selected as the period of rhythm [5]. Lomb-Scargle periodogram is another algorithm for cycle discovery based on Fourier analysis and is mainly used for unequal distance data [53]. Saner et al. assessed blood pressure and heart rate with Fourier analysis and found that cardiovascular rhythmicity is related to obesity in children [126]. Chi-square periodogram uses Chi-square statistics, has been applied to model the period of lobsters'

circadian rhythms to enhance fishing efficiency [61]. Chi-square periodogram calculates the variance between-period and within-period, and use χ^2 distribution to evaluate the significant level. Vukolic et al. applied Chi-squared to model the circadian rhythms of blood pressure and heart rate of mice carrying the muted gene *Per2*, and found the circadian clock gene *Per2* control cardiovascular rhythms [76]. Cyclic Hidden Markov Model (CyHMM) has also been used for modeling cyclic time series data. CyHMM outputs the period’s length by inferring the cyclic latent states of input time series [147]. Pierson et al. applied the cyHMM to sleep time, steps, and calories burned and found that these features showed a weekly cycle. All these statistical models could output the general periods of time series. Still, they all lose the detailed information within each period simultaneously and do not capture the variation in the sequence shape in each cycle [147]. To summarize, the previously discussed approaches have efficiently identified human activity periods by analyzing sensor signals data. However, the cyclic time series modeling approaches mentioned above assume stability in the time series, neglecting the inconsistent nature of human behaviors (e.g., irregular rhythms), which may lead to imprecise outcomes.

Shape Estimation of Cyclic Time Series

Once the period is known, cyclic functions can be applied to the time series data to model the rhythmicity. Cosinor is a standard method to model the amplitude and phase of rhythms when the period is used as input. It uses a linear combination of cosine curves to fit time series data using least squares regression [3]. Doryab et al. used the Cosinor model to extract rhythm features from cancer patients and to predict the readmission probability using these features [152]. BIO_CYCLE based on deep learning is another method to model the shapes of rhythms, which can complete the

step of period estimation and waveform extraction in one way[117]. BIO_CYCLE works by training a CNN (Convolution Neural Network) with a periodic simulated dataset. Then, the trained CNN is used to estimate the period, amplitude, phase, and phase lag of a new series. As a result, BIO_CYCLE achieved state-of-the-art performance on several large synthetic and biological time-series datasets. However, BIO_CYCLE can only model the waveform of the most dominant period, while Cosinorcan model more than one periodic components simultaneously. In summary, numerous valuable rhythmic characteristics can be obtained via the processes of periodicity identification and form approximation. The framework I have developed encompasses stationary rhythm modeling and integrates these extracted features for subsequent analysis. These features can be employed in diverse applications, such as serving as input for machine learning algorithms to generate predictions or conducting similarity assessments between distinct rhythmic models. Moreover, there might be several cyclic patterns occurring within a time frame, with an unknown number of such patterns, but these techniques are limited to modeling the primary or a small number of rhythms.

Wavelet Transform for Non-stationary Time Series

Wavelet transform can effectively capture variations in periods [33]. Therefore, Wavelet transform has emerged as a valuable technique for human behavior analysis due to its ability to handle non-stationary and irregular patterns. This method enables researchers to study time-varying features and periodic patterns at multiple scales, offering a richer representation of the underlying behaviors.

Several studies have successfully utilized Wavelet transform in modeling human behavior. For example, researchers have applied this technique to analyze physiological

signals, such as heart rate variability (HRV), to study the impact of stress on individuals. Additionally, wavelet-based methods have been used to analyze and model physical activity patterns, sleep cycles, and mood fluctuations. Poon et al. used the Morlet continuous Wavelet transform to study ultradian activity rhythms [27]. Price et al. used continuous Wavelet transforms with ridges to analyze cell luminescence data [66]. Etchegaray et al. used a similar Wavelet analysis to analyze the instability of the circadian period in mouse SCN neurons [72].

In these cases, investigators used their eyes to process and understand rhythmic patterns from the output of the Wavelet transform, which is a spectrum image and hard to be processed by statistical and traditional machine learning methods. Analyzing spectrum images by human eyes can deal with small sample datasets; however, it is impossible to process massive multi-modal sensing datasets using human eyes. Therefore, I propose a novel rhythm modeling pipeline based on Wavelet transform and transfer learning to process the massive mobile sensing data automatically. Specifically, the rhythm modeling pipeline first transforms the time series into Wavelet power spectrum images. Then, the transfer learning model assigns similar rhythm patterns with the same category in the spectrum image.

Variability Detection of Cyclic Time Series

Changing point detection (CPD) helps to identify the time when human rhythms have been disrupted. CPD has been used in monitoring cyclic human behaviors widely. For example, applying CPD to heart rate (HR), electrocardiogram (ECG), and electroencephalogram (EEG) has helped better diagnosis of heart disease and understand brain activity [56, 96, 50, 46, 155]. CPD has also been applied to human activity recognition using data from smart home and mobile devices. The changing

points in this context represent the transition of human activity [82, 97, 84]. Selection criteria differ among CPD methods, and some methods are sensitive to the changes of amplitude in the mean, variance, correlation, and spectral density. The cumulative sum (CUSUM) is the most familiar CPD algorithm. CUSUM tracks the shift of local mean, and if the decrease or increase of the mean exceeds the threshold, one change will be identified [132]. For modeling the variations in physiological data during each cycle, I adopt the Automatic Non-stationary Oscillatory Modelling (AutoNOM) to model non-stationary time series with a known period. AutoNOM identifies change points in each cycle and achieves piece-wise fitting [155] using sinusoidal regression models simultaneously. I prefer the CPD technology used in the AutoNOM because it is more sensitive to the change of frequencies of time series, and the AutoNOM can find the best sinusoidal equations to fit the data in each segment [155].

2.2.3 Modeling Human Behavior from Mobile Sensor Data

The rapid development and widespread adoption of mobile devices, such as smartphones and wearable devices, have revolutionized the study of human behavior. Mobile sensing techniques enable the continuous, unobtrusive, and objective measurement of various aspects of human behavior [189, 183, 153].

The study of biobehavioral rhythms also relates to research in understanding human behavior from passive sensing data collected via smartphones and wearable devices. Only few studies have actually used mobile data for understanding the circadian behavior of different chronotypes (*e.g.*, [98, 116, 131]). Abdullah et al. [98] analyzed patterns of phone usage to demonstrate differences in the sleep behavior of early and late chronotypes. In a similar study using the same type of data, they showed the

capability of using mobile data to explore daily cognition and alertness [116, 131] and found that body clock, sleep duration, and coffee intake impact alertness cycles.

Data from smartphones and wearable devices has extensively been used for modeling daily behavior patterns such as movement [100], sleep [101], and physical and social activities [110] to understand their associations with health and wellbeing. For example, Medan et al. [73] found that decreases in call, SMS messaging, Bluetooth-detected contacts, and location entropy (a measure of the popularity of various places) were associated with greater depression. Wang et al. [104] monitored 48 students' behavior data for one semester and demonstrated significant correlations between data from smartphones and students' mental health and educational performance. In addition, Saeb et al. [112] extracted features from GPS location and phone usage data and applied a correlation analysis to capture relationships between features and level of depression. They find that circadian movement (regularity of the 24h cycle of GPS change), normalized entropy (mobility between favorite locations), location variance (GPS mobility independent of location), phone usage features, usage duration, and usage frequency were highly correlated with the depression score. Doryab et al. [150] studied loneliness detection through data mining and machine learning modeling of students' behavior from smartphone and Fitbit data and showed different patterns of behavior related to loneliness, including less time spent off-campus and in different academic facilities as well as less socialization during evening hours on weekdays among students with the high level of loneliness.

Recent tools such as Rhythomic [154] and ARGUS [170] use visualization to analyze human behavior. Rhythomic is an open-source R framework tool for general modeling of human behavior, including circadian rhythms. ARGUS, on the other hand, focuses on visual modeling of deviations in circadian rhythms and measures

their degree of irregularity. Through multiple visualization panes, the tool facilitates the understanding of behavioral rhythms. This work is related to the computational framework for modeling human rhythms. Nonetheless, the majority of research in this area has primarily concentrated on examining typical human rhythmic patterns, such as circadian rhythms, and a narrow range of other human rhythms. Moreover, these methodologies predominantly facilitate the comprehension of rhythms via visual representation, while the proposed framework offers methods for processing various data inputs, extracting information, and identifying and modeling rhythms for each biobehavioral indicator with periods differing from the standard 24-hour cycle. To the best of our knowledge, this constitutes the first computational framework to extract and integrate parameters derived from rhythmic models using a multi-modal mobile sensing dataset.

2.3 Image Representation of Sensor Data

2.3.1 Encode Time Series into Image

In addition to being represented as a timeline function, time series can be transformed into other forms, such as images [77]. Several works have proposed to encode time series into image representations, and these works can be grouped into three categories. The first category is to encode time series into images directly using mathematical models. Specifically, Wang et al. proposed two encoding time series methods. One is based on polar coordinates; the other is based on the Markov Transition process[114, 113, 115]. The approach in[163] uses a similar encoding method but for multiple time series. In addition, Hatami et al. converted 1D time series into 2D images using recur-

rence plot [144]. The second category generates images with features extracted from time series. For example, Guinea et al. extracted variation, maximum, and range of time series, which were then used to generate different regions of an image [190]. The third category transforms series from the time domain to the frequency domain. Approaches have been proposed to use a Short-time Fourier Transform (STFT) to represent the time-series into image form (*e.g.*, [119, 140, 166, 177]). After transforming time series into images, computer vision techniques can be used to classify, predict and cluster time series. There have been many applications of image representation for time series classification ranging from biometric signal analysis (*e.g.*, electroencephalogram (EEG), electrocardiogram (ECG)) in medical or health care field [172, 164, 175, 113] to human activity recognition [190, 173, 171]. In my work, I apply the Wavelet transform on sensing data from multiple sources to capture the rhythmic patterns of human behaviors, which are then visualized on a spectrum image. The Wavelet spectrum image provides an information-rich representation of sensor data, enabling the identification of periods and transition points derived from these images.

2.3.2 Transfer Learning for Image Processing

Transfer learning has a wide application in processing images. Transfer learning uses the knowledge gained from one problem (source domain) and applies it to a different, but related problem (target domain) [128, 75]. Based on the knowledge transferred from the source domain to the target domain, there are three types of learning strategies [70, 176]: 1) inductive transfer learning, where all the instances and their corresponding labels are transferred from the source domain to the target domain; 2) instance transfer learning, where only the instances are transferred; and 3) parameter transfer learning, in addition to the instances and labels, a model

pre-trained on the sources domain is applied on the target domain. There are also three requirements to achieve transfer learning: developing an open-source pre-trained model by a third party, re-purposing the model, and fine-tuning for the new problem. In Image Classification, some large viral datasets are used across research, industry, and hackathons. The following are some of the prominent ones: ImageNet [68], CIFAR [69], MNIST [90]. MNIST is a widely used dataset of handwritten digits that contains 60,000 handwritten digits for training a machine learning model and 10,000 handwritten digits for testing the model. The CIFAR-10 dataset contains low-resolution (32x32) color images in 10 different classes. The ten classes represent airplanes, cars, birds, cats, deer, dogs, frogs, horses, ships, and trucks. There are 6,000 images of each class. ImageNet is the state of the art largest image dataset, containing more than 14 million images, usually for 1000 classes. A pre-trained model is already trained on the same domains. There is a lot of open-source pre-trained architectures are available, including VGG-16 [103], ResNet50 [122], Inceptionv3 [127], and EfficientNet [161]. In my work, I use inductive transfer learning and the ResNet50 neural network to transfer the knowledge of real-world objects in the ImageNet dataset into the patterns shown in the Wavelet power spectrum. This is the first work to apply the transfer learning model to spectrum images, and I will then describe how I verified the feasibility of this experiment.

Chapter 3

CoRhythmo Framework for Modeling Cyclic Human Behaviors

In this thesis, I created a computational framework to model cyclic human behaviors using multi-modal sensor data, as illustrated in Figure 3.1. The framework comprises several components, each focusing on a specific aspect of rhythm analysis. The initial component, data pre-processing, handles the transformation of noisy and incomplete sensor streams into data suitable for subsequent rhythm modeling. The second component, rhythm modeling and analysis, can model rhythms from either stationary or non-stationary time series. A stationary time series exhibits consistent rhythms over time, while a non-stationary time series displays rhythms that change over time, such as varying frequency. Consequently, the methods for analyzing stationary and non-stationary time series differ.

For stationary time series, the framework includes a sub-component called periodicity detection, which estimates periods using a frequency-based model. Once significant periods are identified, a rhythm characteristics modeling component (e.g., Cosinor-model) estimates shape parameters of rhythms (e.g., amplitude and phase). These rhythm features can then be used as input in machine learning algorithms to predict outcomes. Additionally, a novel method based on Pearson correlation analysis is proposed to determine the multidimensional relationships between rhythms and health

and well-being outcomes.

In the context of rhythm variability analysis, a change point detection method is employed to identify abrupt frequency changes within each rhythmic cycle. Moreover, I proposed a rhythm variability measurement method named COSANOVA, which utilizes rhythm parameters from the Cosinormethod and allows for the examination of significant differences between rhythmic components across different populations or time windows.

For modeling rhythms from non-stationary time series, an innovative unsupervised Wavelet Transfer Learning (WTL) method is introduced to identify various rhythmic patterns and assign similar patterns with identical pseudo-labels. The Wavelet transform is incorporated in the WTL to generate a two-dimensional image representation of the sensor signals, which displays the varying frequency content of the signal over time. The pseudo-labels generated by the Wavelet Transfer Learning model aid in identifying sub-patterns of human behaviors.

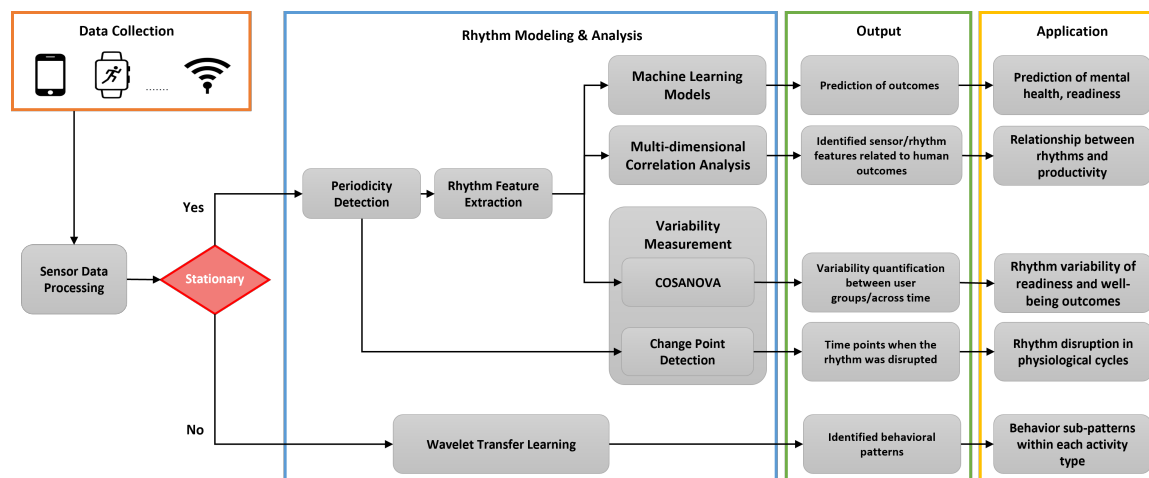


Figure 3.1: Flowchart of the proposed computational framework for the analysis of cyclic human behaviors from multi-modal sensor data.

Finally, I explore different uses of the suggested framework for modeling cyclic human

behavior, and the outcomes of these applications are discussed in Section 4. By applying this framework to a variety of multi-modal mobile sensor datasets, I cover a range of topics such as predicting mental health and readiness, finding the connection between daily patterns and productivity, assessing changes in cycles related to readiness and well-being, identifying interruptions in physiological cycles, and recognizing sub-patterns within each type of activity. This thorough investigation emphasizes the adaptability and potential influence of the framework in various aspects of human behavior research, ultimately demonstrating its effectiveness in tackling real-world challenges.

3.1 Automated Pipeline of Processing Multi-modal Sensor Data for Modeling Human Behaviors



Figure 3.2: Automated pipeline of processing multi-modal sensor data for modeling human rhythms.

To process the massive longitudinal multi-modal mobile sensing datasets and model rhythm with different time granularity automatically, I proposed the pipeline shown in Figure 3.4 [194]. The pipeline incorporates data streams from mobile and wearable devices, including behavioral signals such as movement, audio, Bluetooth, WiFi, GPS, and logs of phone usage and communication (calls and messages); and biosignals such as heart rate and skin temperature, and galvanic skin response. The pipeline performs preprocessing steps such as synchronization, missing data imputation, and

normalization to ensure data quality. Granular features are then extracted to capture biobehavioral patterns such as activity, sleep, social communication, work, and movements. The data streams of biobehavioral sensor features are segmented into different time windows of interest and sent to a rhythm discovery component that applies periodic functions on each windowed stream of the sensor feature to detect their periodicity. The detected periods are then automatically fed into the rhythmic function that represents the time series data stream for that sensor feature. The output of the pipeline is a rhythm model (*e.g.*, the blue line shown in Figure 3.4). The parameters generated by the rhythmic function are used in two ways. First, they are aggregated and further processed to characterize the stability or variation in rhythms over a certain time segment. Second, they are used as features in a machine learning pipeline to predict an outcome of interest (*e.g.*, health status). The two applications of rhythm parameters will be further discussed later. The following sections provide details on the methods used in different components of the framework.

Handling Missing Values

As mobile sensing datasets collected in the wild are expected to include noise and missing data, I developed strategies to handle missing data. The missing values were filled separately for different participants and sensor signals using the local moving average commonly used in time series. For example, if the hourly values of *location* (v) were missing at 2 pm and 3 pm on one day for participant A, then I imputed the values as following: $v_{2pm} = v_{1pm} + (v_{4pm} - v_{1pm})/(4 - 1)$ and $v_{3pm} = v_{1pm} + 2 * (v_{4pm} - v_{1pm})/(4 - 1)$. Moving average is the most suitable interpolation method for rhythm modeling. Other methods such as multiple interpolations and EM estimation introduce cross-correlation between features, and regression estimation

and K-nearest neighbor increase auto-correlation of a single sensor feature [52, 169]. However, the moving average method is sensitive to the number of continuous missing data. If the missing block is large, the moving average will introduce high noise and bias, and the data may need to be removed instead of imputed.

Time Series Segmentation

Windowing is one of the most frequently used processing methods for streams of data. A time series of length L is split into N segments based on certain criteria such as time. This framework allows different ways to segment the time series, including the widely used tumbling windows, which are a series of fixed-sized, non-overlapping and contiguous time intervals. I call each segment a time window (tw) which is a time series of length l , where $l = L/N$. I also add a second segmentation layer to the time series where at each round k and starting point s ($s = 1 \dots N$), I allow to combine a sequence of k consecutive time windows ($k = 1 \dots N$) starting from time window s (tw_s) to generate time series of length k . I call these segments time chunks (tc). For example, in round $k = 1$, the tc_{11} is a time chunk of length one and starting point of tw_1 and tc_{12} is a time chunk of length one and starting point tw_2 whereas for $k = 3$, the tc_{32} is a time chunk of length three and starting point of tw_2 . Time chunks allow flexible modeling of rhythms in different time periods over the length of the time series. Figure 3.3 illustrates the time segmentation process.

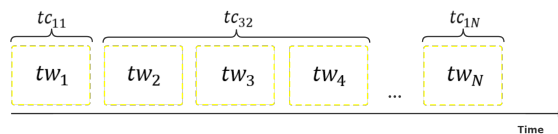


Figure 3.3: The segmentation of time series with time windows (tw) and time chunks (tc)

Feature Extraction

I extracted features in two processing layers. First, I aggregated the raw sensor data into more meaningful behavioral features to capture students' social interaction, physical activity, sleep, and academic life. The collected raw sensor data is just a series of numbers without providing much information. For example, screen data is a time series of values from 0 to 3 (e.g., 0121023...) which does not provide any helpful information, but I can process this time series to extract more meaningful information about how often the user has been interacting with the phone. I then divided each data stream into hourly intervals and extracted behavioral features in each interval following the descriptions documented in [143]. Typical features included statistical measures such as minimum, maximum, mean, standard deviation, length of the status in the hour, and more complex behavioral features such as movement patterns and type and duration of activities. Example features are shown in Table 3.1. Finally, I modeled the cyclic pattern of each behavioral feature using Cosinor, which provided a set of parameters that describe the cyclic pattern. This process and the list of rhythm parameters are detailed in the following section.

3.2 Approaches for Modeling Rhythms from Stationary Time Series

Modeling rhythms from time series data involves analyzing the periodic patterns and extracting meaningful information from them. This process often requires the use of mathematical and statistical techniques to identify and characterize rhythmic patterns in the data. In the following section, I will relevant techniques for identifying

Audio	percentage of time with voice/noise/silence, min/max/mean/std of voice energy
Bluetooth	mean/total number of bluetooth scans
Calories	min/max/mean/total calories burned, min/max/mean/total of decrease in 5-minute calories burned
Location	location variance, percentage of time staying at home, number of visits/time spend at green areas/athletic areas/academic areas/outside campus
Phone Usage	minutes interacting with phone, min/max/mean/std length of interaction periods
Sleep	minutes asleep/awake/restless, min/max/mean length of asleep/awake/restless periods
Steps	total number of steps, min/max/mean/total length of active/sedentary periods
Wifi	number of unique wifi hotspots detected

Table 3.1: Examples of sensor features

periodicity and extracting rhythm features, which are incorporated in the suggested framework.

3.2.1 Periodicity Detection from Stationary Time series

Determining the exact length of a period is crucial for most studies involving rhythms. The frequency-based approach, such as fast Fourier transform (FFT), can find the period value of major rhythms but is limited to stationary time series. Several frequency-based methods for estimating the period have been developed and can be grouped into three main categories: 1) frequency domain methods techniques dependent on periodogram (*e.g.*, FFT) [102]; 2) time domain techniques relying on autocorrelation function (ACF) [105]; 3) state transition techniques utilizing hidden Markov models (HMM) [148]. In this section, several advanced period estimation techniques were employed [188], encompassing Fast Fourier Transform (FFT), Chi-squared periodogram, and CyHMMs. These methods were subsequently applied to a trio of datasets to investigate the potential of identifying human rhythms through mobile

sensor data. The initial dataset, generated synthetically with known periods, was utilized to evaluate the efficacy of the three methods in detecting periods within the data. Following this, the methods were assessed on a pair of real-world mobile sensing datasets; one dataset contained heart rate and temperature data spanning 70 days, while the other dataset comprised 16 days of fine-grained heart rate (HR), heart rate variability (BVP), skin temperature, and galvanic skin response (EDA) collected from the E4 wearable device.

Fast Fourier Transform (FFT)

Fast Fourier transform is an algorithm that converts a signal from the time domain to the frequency domain [5]. In this way, a periodic time series can be expressed by the sum of its frequency components. The Fourier periodogram obtained by the FFT encodes the spectral energy at a given frequency, and the dominant frequency is the component with maximal frequency. The dominant period is the reverse of the dominant frequency.

Chi-squared Periodogram

The Chi-squared Periodogram was developed from the Enright Periodogram [1]. The Enright periodogram is based on the principle that the variances of different segments of the time series are arranged in periodic order sequentially. This process repeatedly divides the long data stream into different periods and calculates a variability index for each period. For the significance test of each hypothesis period, the Enright periodogram uses the F statistic to compare the between-class and within-class variabilities, to test the null hypothesis of the equal class mean. The Chi-square peri-

odogram uses the χ^2 distribution instead of F distribution. Sokolove and Bushell [7] proposed the index Q_P to calculate the significance of different frequencies in time-series data. The stronger is the rhythmicity in a data set, the higher is the value of Q_P . For a dataset with N values (i.e., X_i for $i = 1$ to N), which can be broken down into K sections of period P , the formula of Q_P could be defined as follows [16]:

$$Q_P = \frac{KN \sum_{h=1}^P (M_h - M)^2}{\sum_{i=1}^N (X_i - M)^2} \quad (3.1)$$

where M_h is the mean of P values under each time unit of the period length, and M is the mean of all N values.

Cyclic Hidden Markov Models

Cyclic Hidden Markov Models (CyHMMs) are a special kind of Hidden Markov Models (HMMs) [11] for detecting and modeling cyclic patterns. The input time series will be treated as the observation sequence, and a set of cyclic latent states will be inferred for the observation sequence. The period of latent states will be returned as the period of input time series. Although having the same primary structure as the standard HMMs, CyHMMs differ from HMMs as they do not allow the random transition between hidden states. They require that the transition between hidden states follows a specific order, and the next state of the final state is the starting state, thus forming a closed-loop link to reflect the periodicity. In the CyHMMs, the time spent at a particular stage follows the Poisson distribution, while the standard HMMs have a Geometric distribution [147].

The Appendix A describes the evaluation of the above three periodicity detection methods on three datasets.

3.2.2 Rhythm Characteristic Extraction from Stationary Time Series

Extracting rhythm characteristics is a method used to create representations of the recurring patterns found in human behaviors, as observed through sensor data. These models can come in various forms, such as mathematical equations, as long as they help distinguish different rhythmic features. In rhythm characteristic modeling, most current techniques primarily focus on stationary rhythms. This section intends to give an overview of these stationary rhythm modeling methods. Among the leading approaches, the Cosinormethod has become widely recognized. Furthermore, BIO_CYCLE, a deep learning-based method, has been developed to model rhythm shapes. This technique has the advantage of combining period estimation and waveform extraction in a single process.

Cosinor

The Cosinor method is one of the state-of-the-art methods. The Cosinor model consisting of a linear combination of cosine curves with known periods can be fitted using least square regression [cornelissen2014Cosinor, 2]. After getting the periods of rhythms, the Cosinor method can be used to estimate the shape of rhythms from sensing signals, as shown in Figure 3.4. The black line visualizes the sensing data, and the blue line represents the Cosinor rhythm model. The Cosinor method is mathematically expressed as:

$$y_i = M + \sum_{c=1}^C A_c \cos(2\pi(t_i + \phi_c)/T_c) + e, \quad (3.2)$$

where y_i is the observation at time t_i ; M is the MESOR; t_i is the sampling time; C is the number of input periods; A_c , T_c , ϕ_c represent the amplitude, period, and acrophase respectively; and e is the error. Cosinor also outputs the standard error (SE) for MESOR, amplitude, and acrophase respectively.

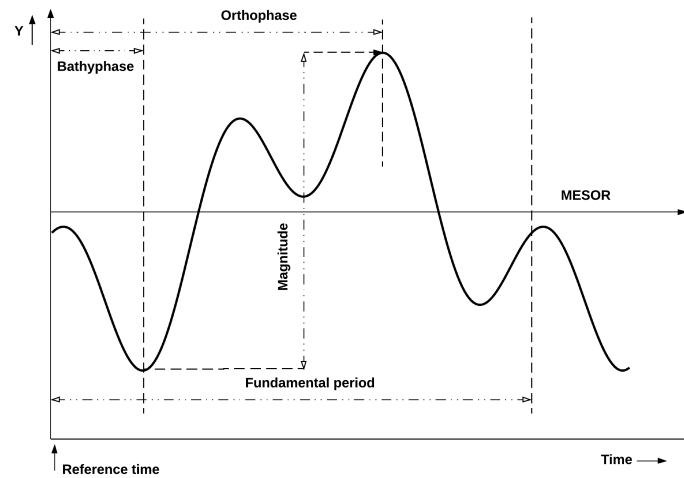


Figure 3.4: A visualization of rhythm features [99].

Table 3.2 provides a complete list of the parameters of the Cosinor rhythm models output by the above pipeline, and Figure 3.4 visualizes these rhythm parameters. The generated rhythm parameters will be used in two ways. First, they are processed and further aggregated to characterize the variation in rhythms. Second, they are used as features in a machine learning pipeline to predict an outcome of interest (*e.g.*, well-being and health status). The processes and results of the usage of rhythm features in rhythm variability and outcome prediction will be described in later sections.

Rhythm Parameters	Definition
Fundamental period	Fundamental period is the least common multiple (LCM) of all individual periods.
MESOR	Estimating midline of the rhythm curves.
Amplitude (Amp)	Half the difference between the maximum and the minimum of the best-fitted curve in an individual period.
Acrophase (PHI)	Lag from a defined reference time point to the maximum point within an individual period.
Magnitude	Half the difference between the maximum and the minimum of the best-fitted curve in the fundamental period.
Bathypphase	Lag from a defined reference time point to the minimum point within an individual period.
Orthophase	Lag from a defined reference time point to the maximum point within the fundamental period.
P-value (P)	P-values indicates the significance level of the model fitted by an individual period.
Percent rhythm (PR)	PR is the coefficient of determination for the model using an individual period.
Integrated p-value (IP)	IP indicates the significance level of the model fitted by the fundamental period.
Integrated percent rhythm (IPR)	IPR is the coefficient of determination for the model using the fundamental period.

Table 3.2: Definitions of rhythm parameters output from the Cosinor model

3.3 A Novel Wavelet Transfer Learning (WTL) Method for Modeling Rhythms from Non-stationary Time Series

In this section, a new Wavelet Transfer Learning Neural Network (WTL) method is developed for non-stationary rhythms from multi-modal cyclic time series. The WTL architecture comprises two main components, as shown in Figure 3.5. The first component employs Wavelet transform (WT) to create an image-based representation of human behavior from raw multi-modal time series data. The images generated by WT undergo a fusion and smoothing process (Figure 3.5) before being passed to the second component. The second component groups the images using a transfer learning Convolutional Neural Network (CNN).

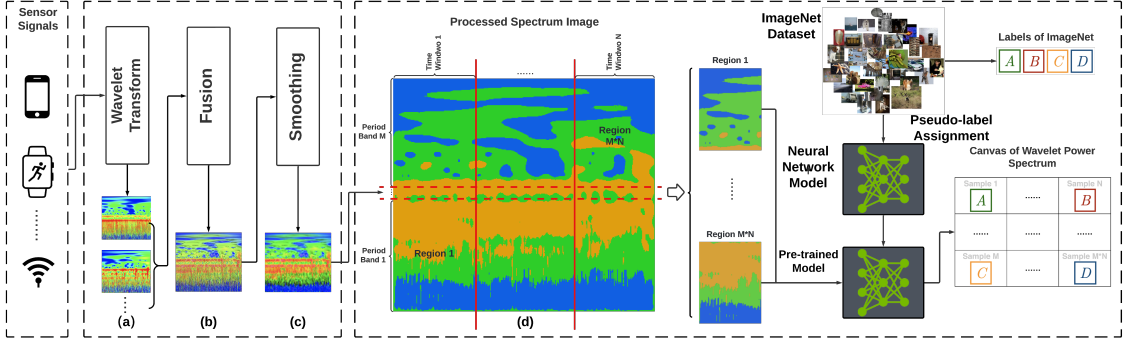


Figure 3.5: The general architecture of Wavelet Transfer Learning (WTL). The Wavelet transform generates a time-frequency power spectrum of single time series sensor signals. The image fusion combines the Wavelet power spectrum images from multiple sources into one. Smoothing makes the fused spectrum images look natural and highlights the significant regions of the spectrum images. The pre-trained neural network model will categorize the smoothed spectrum images into the most similar category from ImageNet and output the corresponding label. (a) spectrum image per sensor signal; (b) fused spectrum image; (c) smoothed spectrum image; (d) reconstructed spectrum image after smoothing.

3.3.1 Multi-modal Time Series Wavelet Transform

Human behavior’s periodic oscillations can also display significant variations in period and amplitude over time. In this section, instances where periods change over time are discussed, and the Wavelet Transform (WT) is employed to provide a power spectrum describing the variability of the periods. WT offers a means of extracting information about time and frequency domains from non-stationary time series, enabling an interpretation of the signal’s variability [159, 19, 71]. WT produces a two-dimensional time-frequency Wavelet power spectrum image, as depicted in Figure 3.6. The X-axis represents the time span of the input time series, the Y-axis denotes the range of frequency expressed in time units, and the rainbow illustrates the power of the distinct frequencies.

The definitions of key elements of the power spectrum in Figure 3.6 are listed as

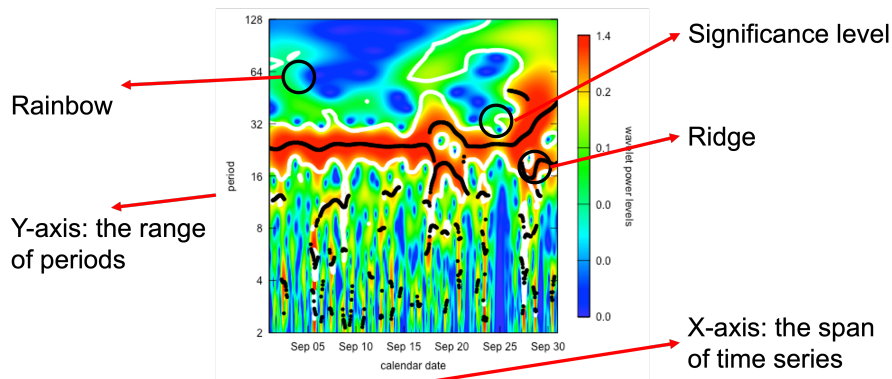


Figure 3.6: An example of Wavelet spectrum image.

follows:

- Y-axis: the range of periods, expressed in time units.
- X-axis: time span of input time series.
- Rainbow: power of distinct period
- Ridge: The Wavelet ridges are the maxima points of the normalized scalogram. They indicate the instantaneous frequencies within the limits of the transform's resolution.
- Significance level: The thick white contour lines represent the significance level and the lighter curve denotes the cone of influence (COI). The area within the white line is the region of significant periods.
- Ridge: The Wavelet ridges are the maximal points of the normalized scalogram. They indicate the most significant instantaneous frequencies at a given time point.
- Cone of Influence (COI): For a time series of finite length, computation of CWT using DFT requires that the time series is cyclic. To satisfy this requirement, the

time series is often padded with zeros. Zero padding leads to errors at the ends of the Wavelet power spectrum. The region in which these edge effects become important is called the cone of influence (COI). As a consequence, the results outside the COI may be unreliable and should be interpreted with caution.

Figure 3.7 shows a comparison between Fourier transform spectrum and Wavelet transform spectrum with the same input time series. The Fourier transform allows us to detect and measure the periods of significant rhythms, but I do not know whether the rhythms have changed over time. However, the Wavelet transform power spectrum includes the rough frequency that happened at the rough time for a signal. As shown in Figure 3.7, 24h is the most dominant period in Fourier spectrum. As for the Wavelet spectrum, I can also observe a 24h rhythm in general, but the rhythm was disrupted at around Sep 20 and had an increase in periods after Sep 25. In addition, Wavelet methods can also be used to assess the strength of signal components lying in different frequency bands. For example, some significant rhythms with periods less than 24h can also be observed, but they did not last for a long time.

These characteristics enable Wavelet analysis to evaluate the power of signals across various frequency bands. I applied the WT to behavioral data, as illustrated in Figure 3.8. The figure displays three examples of Wavelet power spectra obtained from accelerometer data corresponding to three separate human activities. Because the behaviors are different, the three Wavelet spectrograms look significantly different. In Figure 3.8(a), the series is stationary cyclic across time and appears to have two periodic components. The first has a period of around 32 (on the y-axis), and the other around 128. Notice the period of 32 has less power than the period of 128, while the low-period oscillation region (*i.e.*, the area below 16) provides no significant periods. Moreover, the period of 32 is not continuous across time, indicating the regular cyclic

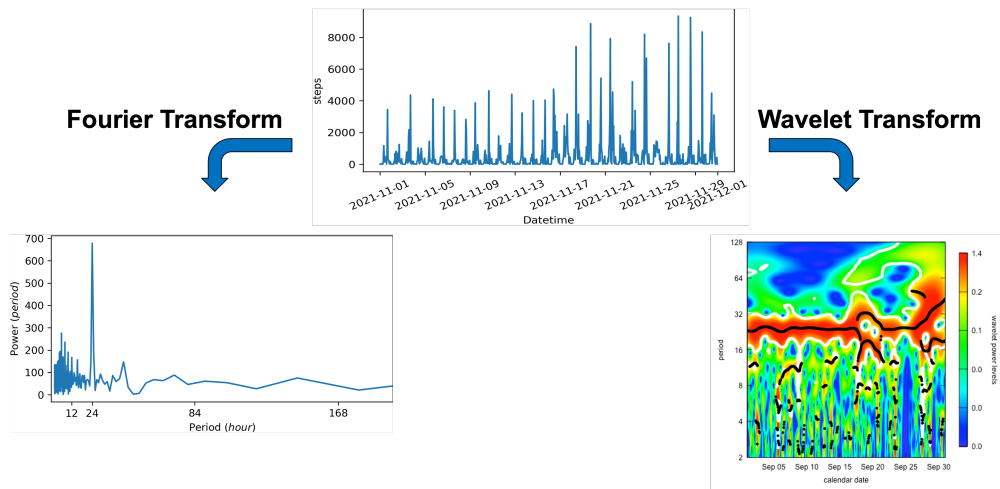


Figure 3.7: Comparison between Fourier transform spectrum and Wavelet transform spectrum with the same input series. The input time series is one person's one month's steps data (sampling frequency: 1h)

behavior is disrupted. The cycling sample presented in Figure 3.8(b) can be periodic but not stationary because its periods change over time. In the green window of the raw sensor data in Figure 3.8(b), a stable period can be observed between 64 and 128. However, in the red window, the red color of the Wavelet power spectrum also appeared in the region above the 128. And in the blue window, the red color mainly concentrates on the region below the 32. This observation indicates that the cycling periods are initially between 64 and 128, then increase in the green window and drop in the blue window. The time series for the standing activity visualized in Figure 3.8(c) is non-cyclic and non-stationary. However, as shown in the black window of Figure 3.8(c), the color and pattern of the Wavelet spectrum changed when the time series fluctuated. This example shows the Wavelet transform can analyze the non-stationary patterns and provide visual information about the characteristics of the time series that otherwise could not be observed.

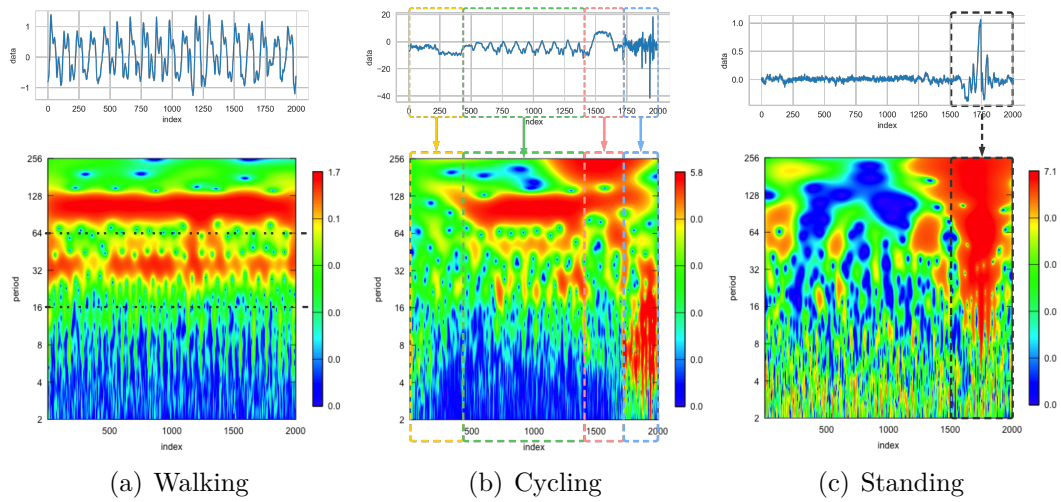


Figure 3.8: Wavelet power spectrum images of accelerometer samples from three human activities. The plot depicts the presence of distinct periods throughout the time series. The time of the respective periods on a logarithmic scale at the ordinate and the power of distinct periods as a function of the time shown in the rainbow colors. The sample shown in Figure 3.8(a) is stationary and periodic. The sample in Figure 3.8(b) is periodic but has changed over time and Figure 3.8(c) is non-periodic. The Wavelet spectrum shows significant differences when the time series fluctuates violently.

Fusion of Single Signal Wavelet Spectrums

Different aspects of behavior are often characterized more accurately by interpreting signals from multiple sensors. The approach can handle multi-modal data by creating one Wavelet image per sensor. To create a single image from multiple WT images, I adopt the fusion strategy introduced in [157] to merge the Wavelet spectrum images (See Figure 3.5). This fusion method can be applied to any number of input images. The technique has been validated on multiple medical fusion categories and achieved state-of-the-art performance in visual quality. The intuition behind the fusion is to gather and preserve common and significant behavioral information from multiple images into a single image. For example, multiple signals may have different frequencies and time-dependent patterns that may or may not be significant for characterizing

the overall behavior. However, if the pattern is preserved in the fusion of the Wavelet spectrum images, it may signal behavior characteristics.

Power Spectrum Smoothing

As shown in Figure 3.5, I smooth the fused Wavelet spectrum images in two steps. First, I perform a convolution operation between the given image and a predefined low-pass filter kernel to generate more natural-looking images. I choose the median filter, which computes the median of all the pixels under the kernel window and replaces the central pixel with this median value.

Next, I smooth the color spectrum of the fused image to obtain fewer and more distinct colors that highlight significant regions of the image. This process partitions the image into multiple distinct regions containing sets of pixels with similar attributes. In this way, the original fused time-frequency spectrum images are transformed into new representations that are more meaningful and easier to analyze. I utilize the K-Means clustering algorithm to segment the interest areas into pieces with different colors. The values of the R, G, and B components in the color of each pixel are input into the K-mean algorithm. Since the Wavelet spectrum comprises the blue, red, and green colors, I set the number of clusters K equal to 3 and are more interested in the area with red indicating more powerful frequencies.

3.3.2 Behavior Modeling and Interpretation via Fused Wavelet Spectrum Transfer Learning

After generating fused and smoothed Wavelet images, I assign pseudo-labels to the images via transfer learning. There are several advantages to this approach. Specif-

ically, transfer learning can bridge between a source domain with rich sample data and a target domain with only a few or even zero labeled samples [128, 75, 167]. This allows the approach to be parameter-free; neither the number of clusters, a distance metric, nor any other hyperparameter must be set. I chose to use ResNet-50 [122] trained on ImageNet dataset [137] to label the input Wavelet power spectrum images. The ImageNet dataset comprises everyday objects such as an “envelope” and “jersey”. Thus, because the label assigned by the network characterizes the predominant shape in the Wavelet image, the properties of the behavior encoded by the Wavelet image can be intuitively understood. I call this assigned class a pseudo-label.

Because the source dataset (ImageNet) significantly differs from the target dataset (the Wavelet images), I use an unsupervised domain adaptation approach to improve the network accuracy. Specifically, I use maximum mean discrepancy (MMD), a widely used non-parametric criterion that measures the discrepancy between the distributions of source and target domains [95]. In the implementation, I used the MMD metrics in the fully connected layer to reduce the difference between the source and target feature domains. I run a small test to confirm that the pseudo-labels align with the spectrum images. I count the number of each type of pseudo-label assigned to the image by WTL. I then choose the three most frequent pseudo-labels — “Envelope,” “Jersey,” and “Chain” — and find the corresponding real-world image from the ImageNet Dataset using the Structural Similarity Index (SSIM) [79]. The SSIM is a common method for quantitatively measuring the similarity between two images. The SSIM values range from 0 to 1, where 1 is a perfect match between the original image and the copy. As shown in Figure 3.9, the SSIM between the spectrum images and ImageNet pictures is relatively high, indicating that the chosen pseudo-labels are close to the real-world representation of the fused spectrum images.

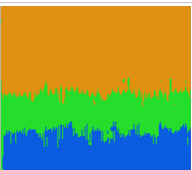



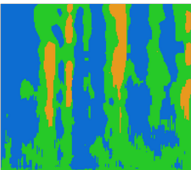
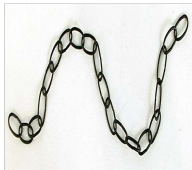
Pseudo Label	Processed Spectrum	ImageNet Picture	SSIM
Envelope			0.81
Jersey			0.73
Chain			0.69

Figure 3.9: The figure shows the three fused spectrum images with their assigned pseudo-labels are close to the real-world representation of the pictures in the ImageNet using the Structural Similarity Index(SSIM) [79].

3.3.3 Granular Behavior Inspection with Power Spectrum Segmentation

The processed spectrum images encode various behavioral information about time and frequency from multiple sensors in different image areas. While the whole image represents the overall behavior, different areas of the image can reveal various behavior characteristics. For example, if the image is split into regions across the x-axis, different behavior time windows can be assessed. Similarly, if the image is split into regions across the y-axis, behavior frequencies can be analyzed — the top region represents relatively long frequencies, while the bottom represents relatively short frequencies. This is an important benefit of the approach because human behaviors can vary significantly across frequency and time.

More specifically, an input image $X \in \mathbb{R}^{c \times h \times w}$ (c , h , w representing the channel,

height, and width) is split into $M \times N$ parts $\text{Region}_1, \text{Region}_2, \dots,$ and $\text{Region}_{M \times N}$, where M represents the number of period oscillation bands, and N represents the number of time windows generating sub-images of size $c \times \frac{h}{M} \times \frac{w}{N}$. The $\text{Region}_{M \times N}$ shows the change of power from the time $\frac{N-1}{N} \times w$ to w for the periods $\frac{M-1}{M} \times h$ to h . Before feeding it into the neural network model, I resize the sub-images to 224×224 (the input size of the transfer learning neural network model). I use all the sub-images as independent samples of the neural network.

3.4 Multi-dimensional Correlation Identification in Rhythm Analysis

In this section, I concentrate on exploring assessing possible connections between rhythmic events and specific human outcomes, such as health and well-being. Gaining insights into these relationships is crucial for enhancing our comprehension of the challenges associated with health and well-being. Initially, the Pearson correlation analysis is employed to identify the relationships between rhythmic patterns and the targeted human outcome across various time periods. It is important to recognize that these connections exhibit a multidimensional nature, incorporating various sensing technologies, feature attributes, and rhythm-related parameters. To accurately quantify this complex, multidimensional relationship, a novel aggregation methodology has been formulated and implemented. First, I calculated the correlation coefficient between each rhythm parameter and productivity score to understand how rhythm parameters correlate with the targeted human outcome and whether the correlation is consistent across time windows. To measure correlation, I first normalized the productivity score and all rhythm parameters using min-max normalization and

then calculated a Pearson's correlation coefficient and a 2-tailed p-value for testing the significance. The first step resulted in one correlation coefficient and one p-value per behavioral feature, per rhythm parameter, and per time window as shown in Figure 3.10 (step 1).

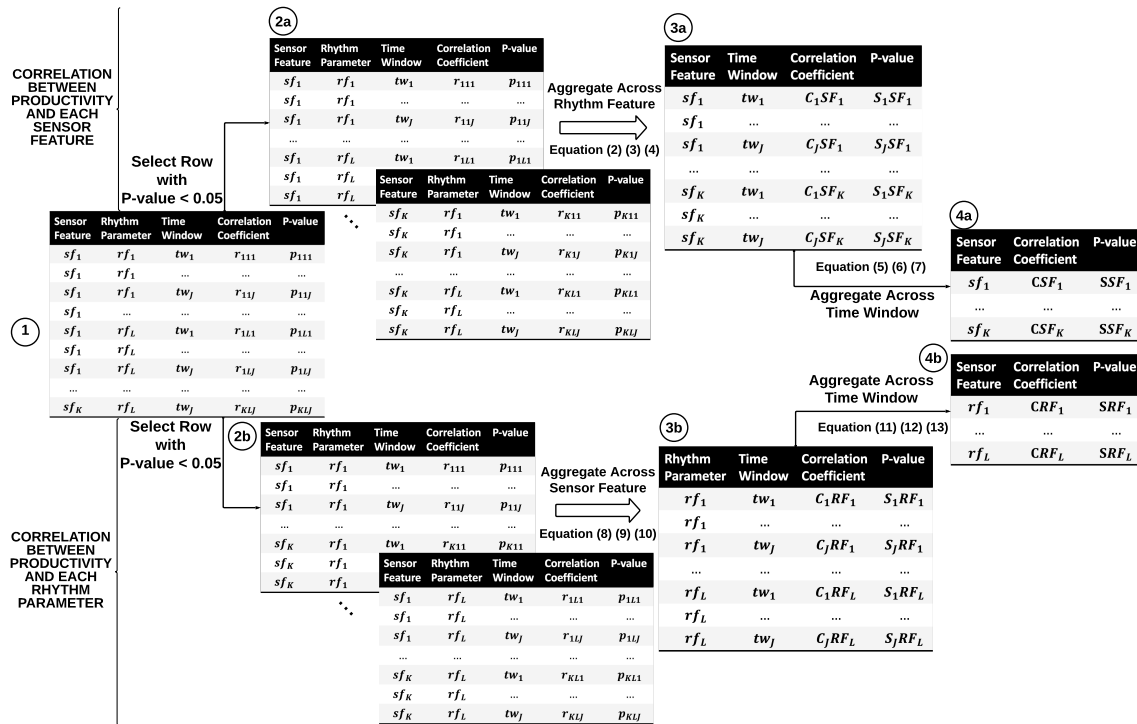


Figure 3.10: The pipeline to aggregate the correlation for a multidimensional dataset with K sensor features, L rhythm parameters, and J time windows. The pipeline is able to output the correlation between productivity and a single sensor, and the correlation between productivity and a single rhythm parameter.

Next, as presented in Figure 3.10, I adopt the Fisher's method to combine the correlation coefficient (r) and its significance (p -value) of every combination of behavioral feature - rhythm parameter - time window. Fisher's method is a widely used meta-analysis technique used for combining the results from several independence tests [15, 179]. These combinations offer insights into the variations in rhythms related to the targeted outcome for each behavioral feature per week (2a in Figure 3.10), as well as

the variations in each rhythm parameter per week related to the targeted outcome (2b in Figure 3.10), irrespective of the specific behavior. While the correlation coefficient represents the strength and direction of the relationship, its significance reflects the reliability and generalizability of the relationship. We, therefore, aggregated significant correlation coefficients for all rhythm parameters per behavioral sensor feature (2a) as well as aggregated significant correlation coefficients for all sensor features per rhythm parameter per week (2b). In step 3 (3a and 3b), I further combined correlation coefficients and significance scores across all three weeks. The final step (4), summarizes the correlation (and significance) values into one final score for each sensor feature (4a) and for each rhythm feature (4b).

Let SF_{lj} denote the set of sensor features that are present for rhythm parameter l in time window j , $SF_{sig,lj}$ denote the subset of SF_{lj} with $p \leq threshold\ t$ (p represents the level of statistical significance and typically ≤ 0.05 , here $t = 0.05$) for rhythm parameter l in time window j . Let RF_{kj} denote the set of rhythm parameters that are present for sensor feature k in time window j , $RF_{sig,kj}$ denote the set of RF whose $p < threshold\ t$ for sensor feature k in time window j .

For each sensor feature sf_k ($k = 1 \dots K$), I first transform the sample correlation r_{klj} to Fisher's z score by Equation 3.3.

$$z_{klj} = 0.5 \times \ln\left(\frac{1 + r_{klj}}{1 - r_{klj}}\right) \quad (3.3)$$

, where r_{klj} represents the significant Pearson correlation coefficient between rhythm parameter l and productivity in time window j for sensor feature SF with the index of k . RF_{sig} represent the set of all rhythm parameters with significant correlation.

Then, I calculate the aggregated correlations and significance between productivity

and rhythm parameters (rf_l , $l = 1 \dots L$) within each time window j , $j = 1 \dots J$ (e.g. week 1) using Equation 3.4 and 3.5.

$$Z_j SF_k = \frac{\sum_{l \in RF_{sig,kj}} |z_{klj}|}{n(l \in RF_{sig,kj})} \quad (3.4)$$

$$S_j SF_k = (1 - 2 \times \sum_{l \in RF_{sig,kj}} \log(p_{klj})) \times \frac{n(l \in RF_{sig,kj})}{n(l)} \quad (3.5)$$

These aggregated correlations and significance are denoted as $Z_j SF_k$ and $S_j SF_k$, which are then averaged across J time windows with Equation 3.6 and 3.7.

$$Z SF_k = \frac{\sum_{1 \leq j \leq J} Z_j SF_k}{J} \quad (3.6)$$

$$S SF_k = \frac{\sum_{1 \leq j \leq J} S_j SF_k}{J} \quad (3.7)$$

, where J represents the total number of time windows.

In the end, the aggregated z score ($Z SF_k$) is transformed into the format of correlation coefficients by Equation 3.8. The final average correlations and significance are denoted as CSF_k and SSF_k respectively.

$$CSF_k = \frac{e^{(2 \times Z SF_k)} - 1}{e^{(2 \times Z SF_k)} + 1} \quad (3.8)$$

Similarly for each rhythm parameter rf_l ($l = 1 \dots L$), I first calculate the aggregated correlations and significance between the targeted outcome and sensor features (sf_k , $k = 1 \dots K$) within each time window j , $j = 1 \dots J$, using Equation 3.9, 3.10 and 3.11.

$$z_{klj} = 0.5 \times \ln\left(\frac{1 + r_{klj}}{1 - r_{klj}}\right) \quad (3.9)$$

$$Z_j RF_l = \frac{\sum_{k \in SF_{sig,lj}} |z_{klj}|}{||SF_{sig,lj}||} \quad (3.10)$$

$$S_j RF_l = (1 - 2 \times \sum_{k \in SF_{sig,lj}} \log(p_{klj})) \times \frac{n(k \in SF_{sig,lj})}{n(k)} \quad (3.11)$$

, where r_{klj} represents the significant Pearson correlation coefficient between rhythm parameter l and the targeted outcome in time window j for sensor feature SF with the index of k . RF_{sig} represent the set of all rhythm parameters with significant correlation.

The aggregated correlations and significance are denoted as $Z_j RF_l$ and $S_j RF_l$ and then averaged across J time windows using Equation 3.12 and 3.13. The final aggregated correlations and significance are denoted as CRF_l and SRF_l .

$$ZRF_l = \frac{\sum_{1 \leq j \leq J} Z_j RF_l}{J} \quad (3.12)$$

$$SRF_l = \frac{\sum_{1 \leq j \leq J} S_j RF_l}{J} \quad (3.13)$$

$$CRF_l = \frac{e^{(2 \times ZRF_l)} - 1}{e^{(2 \times ZRF_l)} + 1} \quad (3.14)$$

, where J represents the total number of time windows.

3.5 Prediction of Outcomes

To explore the potential of using biobehavioral rhythms to predict health and well-being outcomes, a machine learning pipeline (Figure 3.11) was designed to process biobehavioral data, extract cyclic parameters using Cosinor models, and construct machine learning prediction models. These models were built using the rhythmic parameters of each behavioral sensor feature to assess the ability of micro-level cyclic features to predict health and well-being. The pipeline manages missing values in both sensor and rhythm features across different time windows, as described in Section 3.1, selects important rhythm features during the training process, and creates machine learning models for outcome prediction. The following discussion describes the components and configuration of this pipeline.

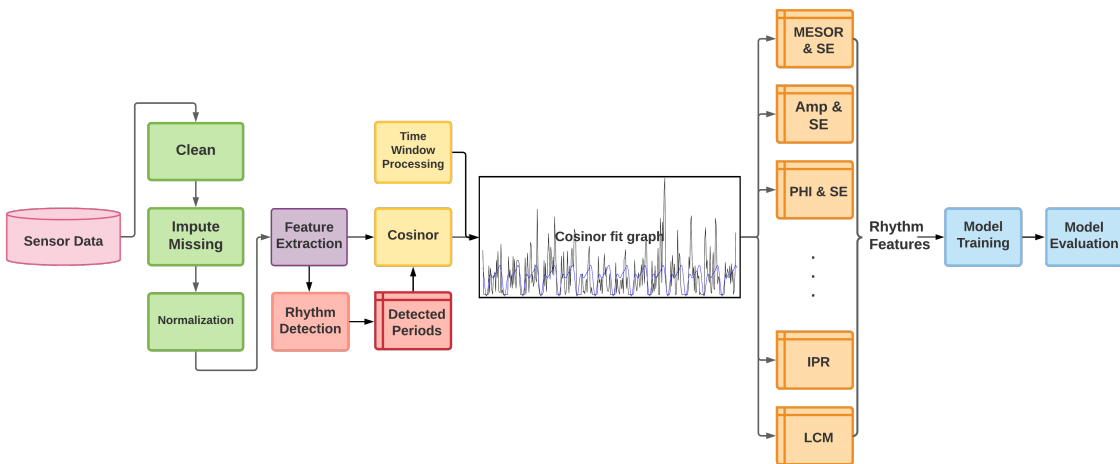


Figure 3.11: Machine Learning Pipeline for Predicting Outcomes via Rhythm Modeling.

3.5.1 Feature Selection

As mentioned in previous sections, for each type of sensor feature, a single period or a multi-frequency Cosinor model is generated which outputs a list of rhythm parameters. These parameters are entered the training process for building machine learning models.

Let M be the number of sensors ($s_1 \dots s_m$), FN_i be the number of features for sensor i and RN_j the corresponding number of rhythmic features for feature j in sensor i . The resulting feature space will be of $M * FN * RN$ which is high dimensional compared to the relatively few data samples for training. As such, a reduction in the number of features is prevalent. The framework allows for the integration of different feature selection methods such as Lasso, Randomized Logistic Regression (RLR), and Information Gain(IG) in the machine learning component.

Lasso is a linear regression model penalized with the L1 norm to fit the coefficients [93]. The Lasso regression prefers solutions with fewer non-zero coefficients and effectively reduces the number of features independent of the target variable. Through cross-validation, the lasso regression can output the importance level for each feature in the training dataset. I use a threshold value of 1e-5 to select features with Lasso, which is the default threshold in the scikit-learn library of Python [85]. Features with importance greater or equal to the threshold are kept, and the rest are discarded.

Randomized Logistic Regression is developed for stability selection of features. The basic idea behind stability selection is to use a base feature selection algorithm like logistic regression to find out which features are important in bootstrap samples of the original dataset [74]. The results on each bootstrap sample are then aggregated

to compute a stability score for each feature in the data. Finally, features with a higher stability score than a threshold are selected. I use 0.25, the default threshold value in the scikit-learn library [85].

Information Gain (also referred to as Mutual Information in feature selection) measures the dependence between the features and the dependent variable (predicted outcome) [48]. Mutual information is always larger than or equal to zero, where the larger the value, the greater the relationship between the two variables. If the calculated result is zero, then the variables are independent. I set our algorithm to select 10 (the default value in the scikit-learn library [85]) features with highest information gain.

3.5.2 Model Building and Validation

The step for building machine learning models using rhythm features of k consecutive time windows and for a population of D data samples is flexible in the framework and can incorporate different supervised and unsupervised machine learning methods such as regression, classification, and clustering. In the current version of the framework, I implement three classification methods, including *Logistic Regression (LR)*, *Random Forest (RF)*, and *Gradient Boosting (GB)*. The choice of algorithms is simply based on our empirical evidence of their performance on this type of data. Logistic regression [44] uses the logistic function to build a classifier. Random forest and Gradient Boosting are two branches of ensemble learning [31] which use the idea of bagging and boosting [35] respectively. Their common feature is to use the decision tree as the basic classifier and to get a robust model by combining multiple weak models. Bagging is short for boost strapped aggregation. Boost strapping is a

repeated sampling method with replacement and random sampling [36]. In boosting, the training set of each iteration is unchanged, but the weight of samples is changed. At each iteration, the training samples with high error rates are given higher weights, so they get more attention in the next round of training.

I built two types of machine learning models: single sensor modeling and multiple sensor modeling. The single sensor model was built with rhythmic features extracted from a single sensor feature alone to better understand the contribution of each sensor feature in prediction. The multiple sensor model on the other hand was used to evaluate the combined power of multiple sensor features. I used a baseline of the majority class to measure the classifiers' performance in predicting the outcome. Again, the flexibility of the framework allows for the incorporation of different baseline measures. Both feature selection process and building machine learning models are done within a cross-validation setting, *e.g.*, leave one sample out [18]. The machine learning component can measure basic performance measures of accuracy, precision, recall, F1, and MCC scores to evaluate the algorithms' performance. From those measures, I choose the results above baseline for each combination of feature selection and learning algorithm to further explore the prediction outcomes.

3.6 Rhythm Variability Modeling

Recognizing and measuring the possible reasons for the period instability in human behaviors is a crucial step in understanding the problems associated with well-being or health. In this section, I initially discuss my efforts to apply change point detection for identifying rhythm disruptions. Subsequently, I introduce a new approach to quantify the variability between rhythm models.

3.6.1 Changing Point Detection in Cycles

Human rhythms can be influenced by external events, leading to inconsistent changes in waveform and impacting the characterization of each rhythm. In this study, after detecting periodicity in the time series data, the focus is on examining the variations within each period to comprehend the potential effects of external factors or fluctuations in every cycle [188]. The Automatic Non-stationary Oscillatory Modelling (AutoNOM) method is employed on biological data to identify change points in each cycle using sinusoidal regression models. The model can be divided into two sub-models: the segment model and the change point model. The rhythm time series y is segmented by k unknown change points by analyzing the frequency change over time. If the frequency on both sides of a time point has a sudden change in the frequency domain, this time point will be regarded as a substation. In each time segment, AutoNOM measures the frequency ω , amplitude β , and phase of the segment σ . ω , β , and σ are multidimensional vectors, and their dimension numbers are unknown. Before using AutoNOM, I need to set the maximum number of change-points k_{max} and the maximum number of frequencies on each segment m_{max} . These threshold setting will transform the problem of modeling rhythm series into a finite state problem. The AutoNOM method will then select an optimal model from the model state space, which is composed of models with a different number of change points and frequencies. The optimal model is the model with the maximum posterior probability, which is calculated as follows [155]:

$$\begin{aligned} \pi(k, m_k, s_k, \theta_k | y) = & \pi(k | y) \pi(m_k | k, y) \pi(s_k | m_k, k, y) \\ & \pi(\theta_k | s_k, k, m_k, y), \end{aligned} \tag{3.15}$$

where k is the number of unknown change-points, m_k is the number of frequency components in each time series segment, s_k is the location of change-points, and θ_k is a three-dimensional vector combined by the frequency ω , amplitude β and phase σ of each segment.

To estimate the parameters k , m_k , s_k and θ_k , a reversible-jump MCMC (Markov chain Monte Carlo) algorithm [21] is applied. The reversible-jump MCMC is an extended version of standard MCMC that provides a simulation of the posterior distribution listed above on spaces of varying dimensions [21]. Thus, the simulation is possible even if the number of parameters in the model is unknown. This algorithm iterates between the segment model move and the change point model move according to its basic structure.

Before applying the AutoNOM, the maximal number of change points k needs to be determined. Although the fitting results will be better as the value of k increases, I want to avoid setting the k too large, because when the value of k is large, minor irregular fluctuations in the time series can falsely identify the existence of change points. I use the mean average percentage error (MAPE) as a measure index to select the AutoNOM model with different input k values. MAPE is a method used to calculate the accuracy of curve fitting and is calculated as follows:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|p_i - a_i|}{a_i} \times 100\%, \quad (3.16)$$

where a_i is the actual value, p_i is the estimated value on minute i and n is the number of minutes for which the data is used.

3.6.2 A Novel Stability Measurement for Human Rhythms from Sensor Streams

As mentioned previously, periodic patterns' stability or deviations from regular ones are important. As such, in this section, I develop a method to quantify variations in rhythms among individual people over different periods (within-person) and across different population groups (between-person). The method employs the population-mean Cosinor method [99] to model a group-level rhythm model for a set of input time series, and an ANOVA significance test [26] is then adopted to identify significant variations. The proposed method is named COSANOVA, which is applicable to the individual biological time series anticipated to be rhythmic with a given period by periodogram method. The Cosinor model is able to be also applied on a group of time series, which is called the population-mean Cosinor model. The ANOVA significance test is used for measuring rhythms stability among different populations through measuring the variance of rhythm parameters obtained from the population-mean Cosinor model.

For each period considered, the Cosinor model gives the estimates of the rhythm parameter MESOR, amplitude, phase, and a standard error for each rhythm parameter. The standard error of the regression, also known as the standard error of the estimate, represents the average distance that the observed values fall from the regression line. Smaller values are better because it indicates that the observations are closer to the fitted line. An ANOVA is used to analyze the difference in the means of different groups (for three or more groups). The ANOVA can be thought of as an extension of the unpaired Student t-test to more than two groups. Or, that is to say, the Student t-test as a special case of the ANOVA for only two groups (or "levels")

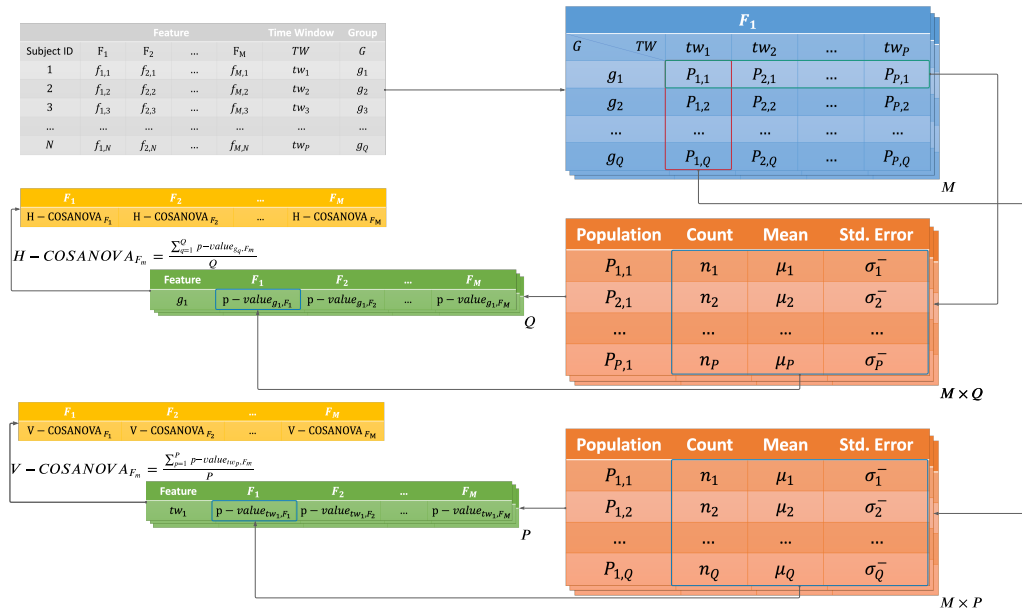


Figure 3.12: The pipeline for measuring rhythm stability. G refers to the group, TW refers to the time window, F refers to the sensor feature, n refers to the number of samples, μ refers to the mean, and σ refers to the standard deviation.

in ANOVA terminology). A two-level ANOVA is algebraically equivalent to a t-test, and produces exactly the same p -values. To use this ANOVA significance test, I have to know how many observations are in each group, and you must know the average (arithmetic mean) and either the standard deviation (SD) or the standard error of the mean (SEM) for the observations in each group. The null hypothesis assumes that there is no variance data in different groups. The following describes the step by step procedure of ANOVA significance.

Step 1: Calculate the Sum of Squares Equation 3.17 is used to calculate the sum of squares between groups (SS_b), and Equation 3.18 is used to calculate the sum of squares within groups (SS_w).

$$SS_b = \sum_{i=1}^k n_i \times std_i \quad (3.17)$$

, where k is the number of groups considered, n_i is the number of subjects in the i -th group, std_i represents the standard deviation (SD) of subjects in the i -th group.

$$SS_w = \sum_{i=1}^k (n_i - 1) \times se_i^2 \quad (3.18)$$

, where se_i represents the standard error of the mean (SEM) of subjects in the i -th group.

Step 2: Calculate the Degrees of Freedom Before proceeding to the next step, I need the degree of freedom. The degree of freedom between groups (DF_b) equals to the number of groups minus one, using Equation 3.19, and the degree of freedom with groups (DF_w) equals to the number of total subjects minus the number of groups, using Equation 3.20.

$$DF_b = k - 1 \quad (3.19)$$

, where k is the number of groups.

$$DF_w = N - k \quad (3.20)$$

, where N is the total number of subjects.

Step 3: Calculate the Mean Squares I now divide the sum of squares by the appropriate number of degrees of freedom in order to obtain the mean squares. The

mean squares between groups (MS_b) and with groups (MS_w) are able to be obtained using Equation 3.21 and 3.22 in the following.

$$MS_b = SS_b/DF_b \quad (3.21)$$

$$MS_w = SS_w/DF_w \quad (3.22)$$

Step 4: Calculate the F-statistic (or F-ratio) The final step of this is to divide the mean square for treatment by the mean square for error using Equation 3.23. This is the F-statistic from the data, and a p -value can be generated from an F – *ratio* score. A p -value less than a threshold (e.g., 0.05) indicates that there is a significant difference somewhere among the various groups; that is, they do not appear to have all come from the same population.

$$F = MS_b/MS_w \quad (3.23)$$

The input to the ANOVA significance test can be the number of samples, standard deviation (SD) and standard error of the mean (SEM) from rhythm parameter MESOR, amplitude, and acrophase for each population across consecutive time windows - Horizontal COSANOVA (HANOVA), and for each time window across different populations - Vertical COSANOVA (VANOVA). In HANOVA, the mean value and the standard error of the rhythm parameters in p consecutive time windows (TW) (tw_1, \dots, tw_p) are used to calculate the significance (p -value) of the variance between the means. In VANOVA, on the other hand, the mean and standard error of the rhythm parameters among Q population groups (G) (g_1, \dots, g_Q) are compared for the

significance of variance in each time window TW . In other words, HANOVA defines the rhythm stability of population level, and VANOVA defines the rhythm stability of time windows. If HANOVA score is greater than the significance level (*e.g.*, 0.05), the group rhythm is stable. Similarly, if VANOVA score is greater than the significance level (*e.g.*, 0.05), the time window rhythm is stable. The scores greater than the significance level mean the variance of rhythm is not significant. The process of calculating HANOVA and VANOVA averages the p -values from ANOVA significance test, as shown in Figure 3.12.

Let $p - value_{g_q}$ be the significance of variance for each rhythm parameter across P time windows for a specific population group g_q . The HANOVA score for a specific sensor feature F_m is calculated as following:

$$H - COSANOVA_{F_m} = \frac{\sum_{q=1}^Q p - value_{g_q, F_m}}{Q} \quad (3.24)$$

, where Q is the number of groups considered.

Let $p - value_{tw_p}$ be the significance of variance for each rhythm parameter across Q population groups for a specific time window tw_p . The VANOVA score for a specific sensor feature F_m is calculated as following:

$$V - COSANOVA_{F_m} = \frac{\sum_{p=1}^P p - value_{tw_p, F_m}}{P} \quad (3.25)$$

, where P is the number of time windows considered.

To assess the reliability of measuring variability in cyclic time series, knowing the ground truth of variability patterns is crucial. As a result, using real-world data for the initial evaluation of the proposed algorithm is not practical. Instead, I create a

simulated population of time series, each consisting of a multivariate time series with added noise to mimic real-world scenarios. The fundamental mathematical signal of the simulated data is based on a harmonic sinusoidal function. The equation for the base periodic function is expressed as follows:

$$F(t) = \sum_{c=1}^C A_c \cos\left(\frac{2\pi t_i}{T_c} + \phi_c\right) \quad (3.26)$$

, where y_i is the simulated value at time t_i ; t_i is the sampling time; C is the set of periodic components; A_c , T_c , and ϕ_c respectively presents the amplitude, period, and phase.

Variations in periodic time series can be summarized into several types such as scaling or shift in period, amplitude, and Mesor, which are shown in the following:

1. Amplitude Scaling: $F_{AS}(t) = \beta_1 * F(t)$
2. Mesor Shift: $F_{MS}(t) = F(t) + \beta_2$
3. Phase Shift: $F_{PT}(t) = F(t + \beta_3)$
4. Hybrid: $F_H(t) = \beta_1 * F(t + \beta_3) + \beta_2$

The amplitude scaling is implemented by multiplying the basic periodic function Equation 3.26 with β . The value of β (β_1 , β_2 and β_3) is generated through the Monte Carlo simulation following a normal distribution $N(\mu, \sigma)$. The Mesor shift is achieved by lifting or dropping the whole simulated time series. The Phase shift is introduced by applying a shift to the sampling time. The hybrid is comprise of all the three above variation types together. Once the samples of time series had been generated, White Gaussian Noise (WGN) is added to simulate the real-world situations. The amount

of variability introduced in the MC simulations is controlled by the value of β , and mean μ and standard deviation of normal distribution N , whose setting is listed in Table.

	Mesor	Amp	Phi
No Difference	0.84 ± 0.08	0.76 ± 0.02	0.71 ± 0.08
Amplitude Scaling	0.05 ± 0.09	$1.02E - 41 \pm 1.78E - 41$	0.78 ± 0.01
Mesor Shift	$4.99E - 21 \pm 9.12E - 21$	0.97 ± 0.02	0.94 ± 0.03
Phase Shift	0.62 ± 0.26	0.33 ± 0.16	$4.12E - 3 \pm 5.39E - 3$
Hybrid	$3.29E - 19 \pm 1.63E - 19$	$6.09E - 38 \pm 2.70E - 38$	$2.48E - 3 \pm 4.11E - 3$

Table 3.3: The result of VANOVA on the synthetic dataset.

Table 3.3 and 3.4 represents the results on the synthetic data. A lower value indicates more variability between two groups of time series. If the value lower than a threshold (e.g., 0.05), the difference between time series is assumed to be significant. The proposed variability measurement method can distinguish the time series with or without altering specific simulation parameters. When examining a specific variation type (e.g., Amplitude Scaling, Mesor Shift, and Phase Shift), the scores of the altered parameters are significantly lower than the others. For the hybrid variation type, all three rhythm parameters have aggregated significance values.

	Mesor	Amp	Phi
No Difference	0.91 ± 0.11	0.78 ± 0.04	0.83 ± 0.07
Amplitude Scaling	0.11 ± 0.09	$1.02E - 17 \pm 2.26E - 17$	0.62 ± 0.02
Mesor Shift	$7.42E - 8 \pm 6.97E - 8$	0.83 ± 0.15	0.43 ± 0.05
Phase Shift	0.41 ± 0.13	0.94 ± 0.18	$7.91E - 2 \pm 3.92E - 2$
Hybrid	$1.37E - 8 \pm 1.77E - 8$	$5.71E - 16 \pm 3.12E - 16$	$3.82E - 2 \pm 1.23E - 2$

Table 3.4: The result of HANOVA on the synthetic dataset.

Chapter 4

Applications of the Framework

4.1 Predicting Human Outcomes Based on Rhythm Modeling

The framework includes machine learning models (outlined in Section 3.5) that use rhythmic features derived from the rhythm modeling process for making predictions. To demonstrate the framework's effectiveness in developing rhythm models from micro- and macro-level sensor features and applying them in prediction tasks, I present two distinct cases. The first case uses data from smartphones and Fitbit devices to examine the connection between biobehavioral rhythms and mental health status. The second case studies long-term biobehavioral rhythms from OURA smart ring data and their capacity to predict readiness. Different analysis approaches are employed to highlight the framework's adaptability in managing various data types and assessing multiple outcomes.

4.1.1 Case1: Predicting Mental Health

The dataset comprises smartphones, Fitbit devices, and survey data collected from 138 first-year undergraduate students at an American university who participated in a health and well-being research study. The dataset was previously used in [150] to detect loneliness among college students. Smartphone data was collected through the AWARE framework [107] and included calls, messages, screen usage, Bluetooth, Wi-Fi, audio, and location. In addition, a Fitbit Flex2 wearable fitness tracker tracked steps, distances, calories burned, and sleep; and survey questions gathered information about physical and mental health including loneliness and depression. The survey data was collected at the beginning and at the end of the semester.

The analysis process was performed in two steps: First, I explored the potential of modeling and detecting rhythmicity in passively collected data from students' mobile and wearable data streams. Then I used the built rhythm models to extract features that were fed into machine learning models to explore the relationship between students' biobehavioral rhythms and their mental health. I aimed to answer the following questions:

1. Can we observe rhythmicity in students' biobehavioral data over the course of the semester? If so, are those rhythms consistent throughout the semester or do they change during different periods?
2. Do we observe any difference in biobehavioral rhythms among students with different health status? If so, do healthy students have more stable rhythms?
3. How accurately can models of biobehavioral rhythms predict mental health status?
4. What are the most important characteristics and rhythmic features that reveal change in health status?

Note that the framework provides the ability to generate a large number of observations on the micro- (sensor feature) and macro-level (sensor), but in this paper, I only focus on observations related to the analysis questions.

Sensor Data Processing

The dataset collected from smartphones and Fitbits consisted of time series data from multiple sensors, including Bluetooth, calls, SMS, Wi-Fi, location, phone usage, steps, and sleep. I grouped this time series data into hourly bins and processed it following the approach in [142] to extract features related to mobility and activity patterns, communication and social interaction, and sleep. Examples of such features include travel distance, sleep efficiency, and movement intensity. I then split the semester data into tumbling cyclic time windows of 14 days or two weeks based on empirical evaluation of different lengths of time windows. The university semester in the studied population was roughly 16 weeks long, which could be divided into eight time windows of two weeks except for the last time window that contained only ten days of data (Figure 4.1). I built a model of rhythm for each student and for each time window.

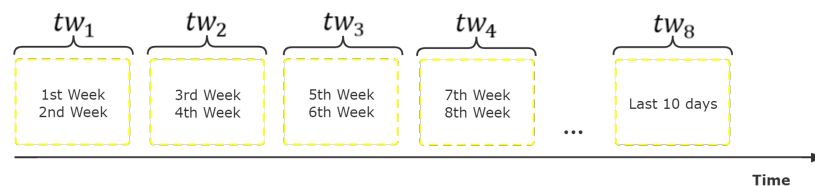


Figure 4.1: The size of a time window is 2 weeks which segments the semester into roughly 8 time windows.

Ground Truth Measures for Loneliness and Depression

In the evaluation, I focused on two mental health outcomes, namely depression and loneliness. These two measures were chosen because of their longitudinal aspect, *i.e.*, lasting for at least a few weeks to enable the investigation of 1) how biobehavioral rhythms of students with mental health conditions would differ from other students and 2) how accurately the state of those mental health conditions could be predicted from extracted rhythms.

Loneliness data was collected using the UCLA Loneliness Scale, a well-validated and commonly used measure of general feelings of loneliness [23]. The questionnaire contains 20 questions about feeling lonely and isolated using a scale of 1 (never) to 4 (always). The total loneliness scores range from 20 to 80, with higher scores indicating higher levels of loneliness. As there is no standard cutoff for loneliness scores in the literature, I followed the same approach in [150] to divide the UCLA scores into two categories where the scores of 40 and below were categorized as *'low loneliness'*, and the scores above 40 were categorized as *'high loneliness'*.

Depression was assessed using the Beck Depression Inventory-II (BDI-II) [22, 28], a widely used psychometric test for measuring the severity of depressive symptoms that have been validated for college students [28]. The BDI-II contains 21 questions, with each answer being scored on a scale of 0-3 where higher scores indicate more severe depressive symptoms. For college students, the cut-offs on this scale are 0-13 (no or minimal depression), 14-19 (mild depression), 20-28 (moderate depression) and 29-63 (severe depression) [28]. For simplicity and to be consistent with the loneliness categorization, I divided these scores into two categories where the BDI-II scores < 14 were labeled as *'not having depression'* and all BDI-II scores ≥ 14 were labeled as

'having depression'.

The machine learning pipeline used these loneliness and depression categories as ground truth labels to classify students' depression and loneliness levels using rhythmic features. Each student filled out the surveys both at the beginning (Pre) and the end of the semester (Post). To capture relationships between biobehavioral rhythms and changes in students' mental health, I categorized students into five groups according to the survey measures for depression and loneliness. For simplicity of representation, I further label *low loneliness* and *no depression* categories as 1, and *high loneliness* and *high depression* as 2. The five mental health categories are as follows:

- All students
- Pre1_ Post1: not having a mental health condition in both pre-semester and post-semester surveys
- Pre1_ Post2: not having a mental health condition in the pre-semester survey, but having it in the post-semester survey
- Pre2_ Post2: having a mental health condition in both surveys
- Pre2_ Post1: having a mental health condition in the pre-semester survey, but not in the post-semester survey

The following paragraphs describe the observations and findings. To distinguish the mental health groups in the two conditions, I add an *L* and *D* to the mental health group for loneliness (*e.g.*, L_Pre1_Post2) and depression (*e.g.*, D_Pre1_Post2) respectively.

Detection of Rhythmicity and Regularity in Student Data

To investigate whether rhythmicity exists in data collected from students' smartphones and Fitbits (Question 1) and whether students' rhythms remain stable throughout the semester (Question 2), I used Autocorrelation and Fourier Periodogram to model students' rhythms in each time window for each sensor feature.

I first applied the Autocorrelation on a sleep feature which indicates that students with high loneliness have less stable sleep rhythms. Figure 4.2 shows the correlogram of the number of restless sleep bouts in two students from different groups, one with low loneliness throughout the semester and the other with high loneliness at the end of the semester. The figure visually depicts differences in the rhythms of these two students where the correlogram belonging to the student with high loneliness projects a less stable rhythm towards the end of the time series. To further quantify such differences in cyclic rhythms of students, I applied Periodogram to 1) detect dominant periods in students' data and 2) measure variability in those periods among students with different health statuses.

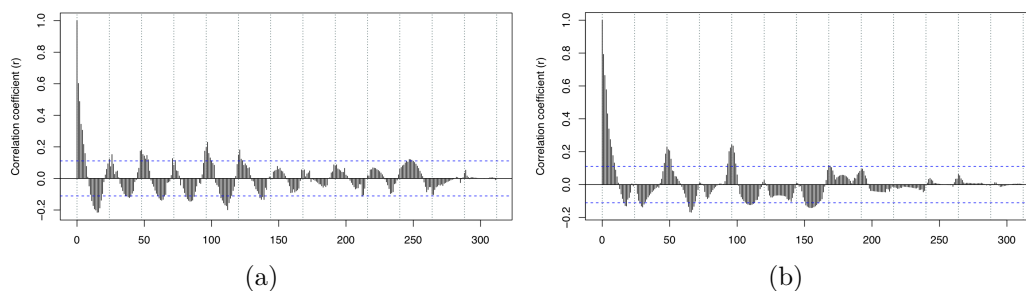


Figure 4.2: Correlograms of feature `num_restless_bout` (number of restless periods in sleep) in time window 4 for two students (left: a student in `L_Pre1_Post1`, right: a student in `L_Pre1_Post2`).

To identify the dominant periods, the Fourier periodogram is used to detect all significant periods for each sensor feature. The results of the periodogram show that

the most dominant cyclic periods in each time window are 24- and 12-hours for all sensor features. For example, for sleep duration feature in the depression category, this trend is consistent in all students regardless of the mental health condition where on average 97.6% and 69.6% of students have 24- and 12-hours as dominant periods in their data across time windows (Tables 4.2 and 4.1). The percentages, however, have a declining trend starting from TW4 (around midterms) towards the end of the semester. This trend can be expected because of the increase in students' workload that causes irregularity in sleep duration. The lowest percentages across all time windows (46.3% on average) are observed in the 12-hour period of students in group D_Pre2_Post2, i.e., students who were depressed throughout the semester. In particular, there is no 12-hour period observed for this group in TW1 (the first two weeks) and TW8 (the last two weeks). The 12-hour or half-day period relates to diurnal/nocturnal activities, and this trend may be indicative of higher irregularity in sleep behavior among students with depression throughout the semester especially at the beginning and towards the end of the semester. The observations are consistent with other studies. [184] observed that older adults with depression have a lower sleep regularity index in a study of 138 participants. [8] observed that irregular sleepers showed more negative moods, including depression, in a study of male college students.

I picked the sleep duration to further analyze changes in periodicity in students who started the semester with normal health status but developed depression or loneliness towards the end (D_Pre1_Post2 or L_Pre1_Post2). Table 4.1 shows that the dominant periods of 24- and 12-hours are preserved for the sleep duration feature in all time windows for both loneliness and depression groups. While the same declining trend towards the end of the semester exists for both loneliness and depression

Pre1_Post2									
Time Window	Loneliness				Depression				
	N	P1 (%)	P2 (%)	P3 (%)	N	P1 (%)	P2 (%)	P3 (%)	
TW1	17	24 (100)	12 (71)	312 (35)	35	24 (100)	12 (89)	312 (34)	
TW2	15	24 (93)	12 (87)	312 (40)	34	24 (97)	12 (88)	312 (38)	
TW3	16	24 (100)	12 (88)	156 (31)	35	24 (91)	12 (80)	156 (31)	
TW4	15	24 (73)	12 (53)	312 (33)	33	24 (91)	12 (64)	78 (40)	
TW5	14	24 (100)	12 (64)	156 (29)	33	24 (97)	12 (58)	312 (36)	
TW6	12	24 (92)	12 (67)	78 (33)	33	24 (94)	12 (64)	78 (45)	
TW7	13	24 (85)	12 (54)	156 (31)	33	24 (91)	12 (61)	156 (40)	
TW8	11	24 (91)	12 (55)	72 (45)	28	24 (93)	12 (78)	72 (32)	

Table 4.1: Top three dominant periods of sleep duration (minutes asleep) feature for Pre1_Post2 groups. N is the number of students in the group. P1 is the most dominant period (*i.e.*, the percentage of students that have this period is highest among all periods). The percentage in parenthesis is the percentage of students that have the period. P2 and P3 are the second and third dominant periods.

groups, a sharper slope is observed for the 12-hour period. The lowest percentage of students in this group with 24- and 12-hour periods are in time windows 4 and 5 with 73% in loneliness category (24-hour), 91% in depression category (24-hour), 53% in loneliness category (12-hour), and 57% in depression category (12-hour). Given that time windows 4 and 5 intersect with midterm and spring break, these observations point to changes in sleep patterns among students whose mental health worsens over the semester.

Group	TW1			TW2			TW3			TW4			TW5			TW6			TW7			TW8		
	N	P1 (%)	P2 (%)	N	P1 (%)	P2 (%)	N	P1 (%)	P2 (%)	N	P1 (%)	P2 (%)	N	P1 (%)	P2 (%)	N	P1 (%)	P2 (%)	N	P1 (%)	P2 (%)	N	P1 (%)	P2 (%)
All	125	24 (98)	12 (70)	120	24 (98)	12 (75)	118	24 (95)	12 (71)	115	24 (88)	12 (51)	104	24 (88)	12 (52)	103	24 (88)	12 (65)	101	24 (89)	12 (53)	97	24 (94)	12 (69)
D_Pre1_Post1	72	24 (97)	12 (69)	68	24 (99)	12 (72)	66	24 (98)	12 (74)	67	24 (85)	12 (46)	60	24 (87)	12 (52)	58	24 (90)	12 (66)	57	24 (88)	12 (51)	58	24 (98)	12 (72)
D_Pre1_Post2	35	24 (100)	12 (89)	34	24 (97)	12 (88)	35	24 (91)	12 (80)	33	24 (91)	12 (64)	33	24 (97)	12 (58)	33	24 (94)	12 (64)	33	24 (91)	12 (61)	28	24 (93)	12 (79)
D_Pre2_Post1	2	24 (100)	12 (100)	2	24 (100)	12 (100)	2	24 (100)	12 (50)	2	24 (100)	12 (100)	1	24 (100)	31.2 (100)	1	24 (100)	12 (100)	1	24 (100)	12 (100)	2	24 (100)	12 (50)
D_Pre2_Post2	16	24 (94)	156 (38)	16	24 (94)	12 (56)	15	24 (87)	12 (40)	13	24 (92)	12 (38)	10	24 (70)	12 (40)	11	24 (64)	12 (64)	10	24 (90)	12 (40)	9	24 (67)	54 (33)

Table 4.2: Top two dominant periods of sleep duration feature for depression groups. N is the number of students in the group. P1 is the most dominant period (*i.e.*, the percentage of students that have the period is highest among all periods). The percentage in parenthesis is the percentage of students with that period. P2 is the second dominant period.

The third dominant periods for sleep duration across all time windows include 312-hour (13 days), 156-hour (6.5 days), and 78-hour (3.25 days). This is an interesting observation as these numbers are multiples of the 78-hour period. In other words,

it seems the sleep duration of roughly one third of the population in these groups follows a weekly pattern that may be imposed by class schedules.

% of Participants with 24-hour period										
Audio	Battery	Bluetooth	Calorie	Location	Location Map	Call&Messages	Screen	Sleep	Steps	Wifi
62	13	42	92	41	17	18	36	69	95	83

Table 4.3: Percentage of participants with 24-hour period across all sensor features

Overall and across all sensor features, I observe the 24-hour as the dominant period for over 52% of the student population with the highest percentages belonging to steps (95%), calories (92%), wifi (83%), and sleep (68%). Table 4.3 presents the overall percentages for each sensor. Calories and steps relate to physical activity. The high percentage of students with 24-hour cycles in these two sensor categories is indicative of regular daily exercise and movement. While there is a low percentage of students with regularity in their cyclic location patterns and visited places (Location Map features), it seems a large number of students have regular daily patterns of using Wifi. This pattern could be expected given that the first-year students live in dorms and are mostly on campus. Interestingly, a low percentage of students seem to have regular cyclic patterns of phone usage (Screen, 36%; Call & Messages, 18%; Battery 13%). While phone use especially battery charging patterns are expected to be cyclic (e.g., charging the phone at night), these observations present the possibility of different phone use behavior among students.

To measure the variability of the dominant periods among students with different health statuses, I look at the percentage of participants in each mental health group that had 24-hours as one of their dominant rhythms for each *time chunk*. This would help observe the extent to which students preserved their normal circadian rhythm over the semester. Recall that time chunks consist of k consecutive time windows, there were 36 different time chunks in total for eight time windows of length 2 in the

dataset. In each time chunk, a participant had 24-hour as a dominant rhythm if and only if this participant had 24-hour as a dominant rhythm in all time windows in that time chunk. Figure 4.3 shows the percentage of participants with 24-hour as the dominant rhythm (y-axis) in each mental health group for each time chunk of length 3 (x-axis). I chose one representative feature from each sensor stream, i.e., Bluetooth (abbreviated as blue in the figure), location (loc), sleep (slp), calories (calor), screen, and steps for further analysis. As shown in Figure 4.3, the trend in the percentage of 24-hour rhythms varies a lot in mental health groups and across time chunks in each sub-figure. To understand the significance of these variations, I 1) applied K-W ANOVA (Kruskal–Wallis one-way analysis of variance) [24] to test the variance of trends across mental health groups, and 2) calculated the variance in the percentage of 24-hour rhythms for each mental health group across time chunks. For loneliness, the trends for all features show significant differences among mental health groups (The average/median of p-value across sensor features is 0.02/0.03). For depression, mental health groups have more similar trends. In contrast to Bluetooth, calorie, and step features that have significant differences in their trends (p-values of 0.05, 0.001, and 0.001), location, sleep, and screen features do not show any significant differences (p-values 0.94, 0.26, and 0.67). This is visually demonstrated in Figure 4.3, e.g., the trend for location is similar for all the depression groups. I also calculated the average variance for each mental health group across sensor features. As shown in Table 4.4 for loneliness, most changes in the 24-hour rhythms were observed in the group with high loneliness at the beginning and low loneliness at the end of the semester (pre2_pre1) group whereas for depression, the group with depression throughout the semester (pre2_pre2) had the largest fluctuations.

For loneliness, the group with low loneliness at the beginning and high loneliness at

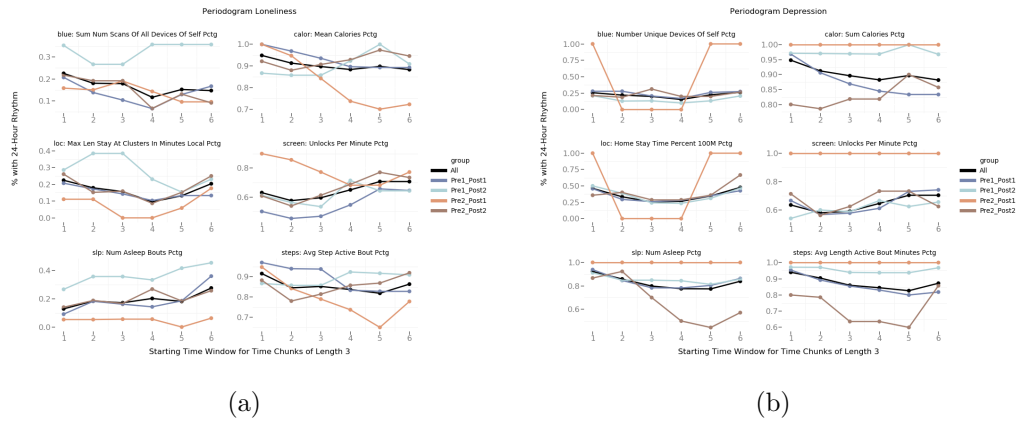


Figure 4.3: The plots show the percentage of participants with 24-hour as the dominant rhythm (y-axis) in each mental health group (left: loneliness, right: depression) for each time chunk of length 3 (x-axis). The data point at $x = i$ corresponds to the time chunk of length 3 starting at tw_i (i.e., tc_{3i}). It represents the percentage of participants with 24-hour as the dominant rhythm in all the 3 time windows tw_i , tw_{i+1} , tw_{i+2} .

the end of the semester (L_Pre1_Post2) shows an overall higher percentage of 24-hour rhythms for features of sleep, location, and Bluetooth across time windows. The opposite group with high loneliness at the beginning and low loneliness at the end of the semester (L_Pre2_Post1) shows a lower percentage of 24-hour rhythms for features of calories and steps but higher percentages for screen features. The Bluetooth feature in the top left of Figure 4.3 (a) which represents the cyclic patterns of the scanned devices belonging to the person is a proxy of social isolation, i.e., the person not being around other people (and their devices) and being mostly by themselves. Starting from TW3 (week 3, 4 and 5), the percentage of students with regular daily cycle for this features in L_Pre1_Post2 and L_Pre2_Post1 groups sharply increase and decrease respectively. In other words, while more students with low loneliness at the beginning and high loneliness at the end of the semester start having a regular social isolation pattern on a daily basis towards the end of the semester, fewer students in the opposite group with high loneliness at the beginning and low lone-

liness at the end of the semester experience this trend. A very similar pattern is observed for another socially relevant feature namely the length of stay in significant locations. The trend is relatively stable and slightly decreasing in students with no loneliness which reflects the stability of behavior in this group. For sleep, steps and, calorie burn, I observe an almost counterintuitive opposite cyclic behavior among L_Pre1_Post2 and L_Pre2_Post1 groups. It seems more students with loneliness toward the end of the semester engage in regular physical activities as projected by calories and steps features and have more regular sleep duration cycles. A relatively similar behavior is observed for the burned calories feature in depression groups (Figure 4.3 top right). While regularity in physical activities slightly increases in students with depression (D_Pre2_Post2), it appears to decrease in students with no depression (D_Pre1_Post1) across time windows. While existing studies, e.g., [67, 92, 150] point to negative associations of physical activities and mental health, I believe the increase in regular physical activities towards the end of the semester may be a coping attempt by students with mental health problems.

Mental Group	all	pre1_pre1	pre1_pre2	pre2_pre1	pre2_pre2
Loneliness	0.04	0.05	0.06	0.07	0.05
Depression	0.05	0.05	0.05	-	0.09

Table 4.4: The table lists the aggregated variance of the percentage of 24-hour rhythms across time chunks for loneliness and depression separately. I first calculated the variance per mental health group in each sensor feature shown in Figure 4.3, and then averaged these variance values across sensor features of loneliness or depression. The aggregated variance can represent the stability of rhythms of each mental health group.

But trends generally look different for depression groups in Figure 4.3 (b). All groups except D_Pre2_Post1 had similar percentages of regular 24- and 12-hour periods for Bluetooth, location, and screen across time windows. While the group with no depression at the beginning and with depression at the end of the semester (D_Pre1_Post2)

shows the highest percentage of normal 24-hour rhythms for features of calories and steps across all time windows, the group that was depressed throughout the semester (D_Pre2_Post2) shows lowest percentages for steps, sleep and calories. In particular, the regularity of sleep in these students seems to decline drastically across time windows. Although expected, this sharp trend is a valuable observation for further exploration of relationships between change in sleep cycles and depression status. The previous study in [184] also observed that sleep irregularity is indicative of depression, but no existing study has analyzed the relationship between change in sleep cycles and change in depression status. The observations provide new findings and insights that call for further and more rigorous investigations.

Prediction of Mental Health Status with Rhythmic Features

The third and fourth questions in the analysis relate to the feasibility of using biobehavioral rhythm parameters to predict students' mental health status. In the framework, I utilize dominant periods that were detected using Fourier Periodogram described in Section 4.1.3 to build Cosinor models of biobehavioral data. This process generates rhythmic features fed into the machine learning process to classify post-semester loneliness and depression categories (low loneliness vs. high loneliness and no depression vs. with depression) of the students. I build two types of datasets, one with single sensors only and one with multiple sensors. In the following paragraph, I will evaluate the performance of single sensor modeling and multiple sensor modeling to find out what types of sensor features and rhythmic features contribute most to the prediction.

For *Single Sensor* datasets, I use the rhythmic features of each sensor feature separately, *i.e.*, for each sensor feature and each time window (with time windows of

two weeks), I take the rhythmic features of this sensor feature and time window to form the input dataset. I remove datasets with more than 30% missing instances (80 training instances) as I consider it too small to generate a reliable and generalizable model. For *Multiple Sensors* datasets, I select the sensor features that provide accuracy above baseline in models built with single sensors. For both approaches, I use the majority class ratio *i.e.*, the category that has the highest percentage of labels for that category as the comparison baseline. I then repeat the same process I followed for single sensor datasets, but this time for the combination of sensor features, *i.e.*, for each combination of sensor features and each time window, I take the rhythmic features of the selected sensor features of those sensors and time window to form the input datasets. Other than the difference in the input dataset, the machine learning pipeline is the same for the two types of datasets.

Given the imbalanced datasets for both health conditions *i.e.*, the different number of samples in the two classes (*e.g.*, 59% of samples in category 1 vs. 41% in category 2 of depression), using the accuracy will not be adequate for performance evaluation and needs to be accompanied by other measures such as F1. For every combination of time window and sensor, the F1 score is used to select the model with the best performance. I build models with single sensor and multiple sensors datasets for both mental health conditions. The results of all combinations are shown in Figures 4.4 and 4.5. The heatmaps use the depth of color to represent the F1 score. Given a large number of features, I only report results with accuracy above the baseline (majority class percentage). Through the single sensor modeling, I can judge which type of sensor is most effective in predicting mental health. Overall, I find that the models with multiple sensors improve the prediction performance. A summarization of the results are listed in Table 4.5.

Single Sensor Modeling The F1 scores of machine learning models with single sensor features are shown in Figure 4.4. Overall, the models for loneliness prediction obtain higher accuracy (F1) scores than depression models (Table 4.5) which may be due to more sparsity in depression datasets. Rhythm parameters obtained from Cosinor models built for features related to Bluetooth, calories, location, sleep, and steps perform better in predicting both loneliness and depression levels. Although the best model to classify post-semester loneliness is built using Gradient Boosting on rhythm parameters of calorie data from tw_1 to tw_3 with an F1 score of 0.76, more models built on rhythms of location and locationMap provide high performance. The best model for post-semester depression with an F1 score of 0.7 is also built using Gradient Boosting but on the locationMap data from tw_3 to tw_5 . Compared to other sensors, models using rhythmic parameters from locationMap features show better performance for predicting post-semester depression (six out of ten models with the highest F1 score use locationMap features). Although the F1 scores of models with a single time window are generally lower than models with multiple time windows, there are some exceptions in the heatmaps of both loneliness and depression. For example, the loneliness model using sleep features in tw_1 achieves an F1 score of 0.75, and the F1 score of the depression model using sleep features in tw_5 equals 0.68. Interestingly and somewhat counter-intuitively, across all sensors, the majority of models (avg. 57.5% for single sensors and 53.5% for multiple sensors) using early semester time windows (tw_1 to tw_4) appear to have higher F1 scores for post-semester loneliness and depression prediction than late semester time windows. I believe this observation provides initial evidence for the possibility of early detection of mental health status via monitoring of changes in biobehavioral rhythms.

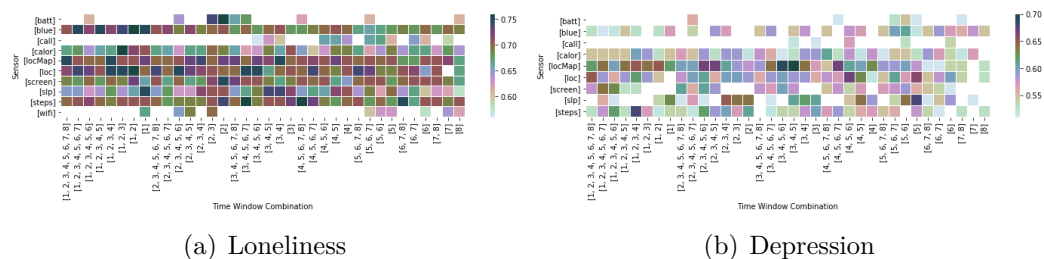


Figure 4.4: The heatmap displays the largest F1 score in the loneliness and depression prediction model trained by a combination of different single sensor features and time windows.

Multiple Sensor Modeling I do the same analysis for the combination of sensor features. From Figure 4.5, I observe that the combination of multiple sensor features contributes to the improvement of the F1 score. For example, the combinations related to steps, sleep, location, calorie, and Bluetooth end with better results. For predicting loneliness, the best model is built with Logistic Regression, which uses the Bluetooth and steps data from tw_5 to tw_8 and obtains an F1 score of 0.91. For predicting depression, the best model is obtained from Logistic Regression using the rhythm parameters from Bluetooth, calorie, location, screen, and steps features. The model only uses tw_6 to predict depression with an F1 score of 0.89. The best model predicting depression has a lower F1 score than the best model predicting loneliness, which is the same as the single sensor model and may be due to sparsity in sensor data.

Table 4.5 summarizes the mean and max of F1 scores for models built with each combination of the feature selection and machine learning methods. In single sensor modeling, the combinations of Logistic Regression with Lasso and Randomized Logistic Regression) perform best for predicting loneliness with the mean and max F1 score of 0.7 and 0.76 respectively. The combination of Gradient Boosting and Information Gain provides the highest F1 score for the prediction of depression. For

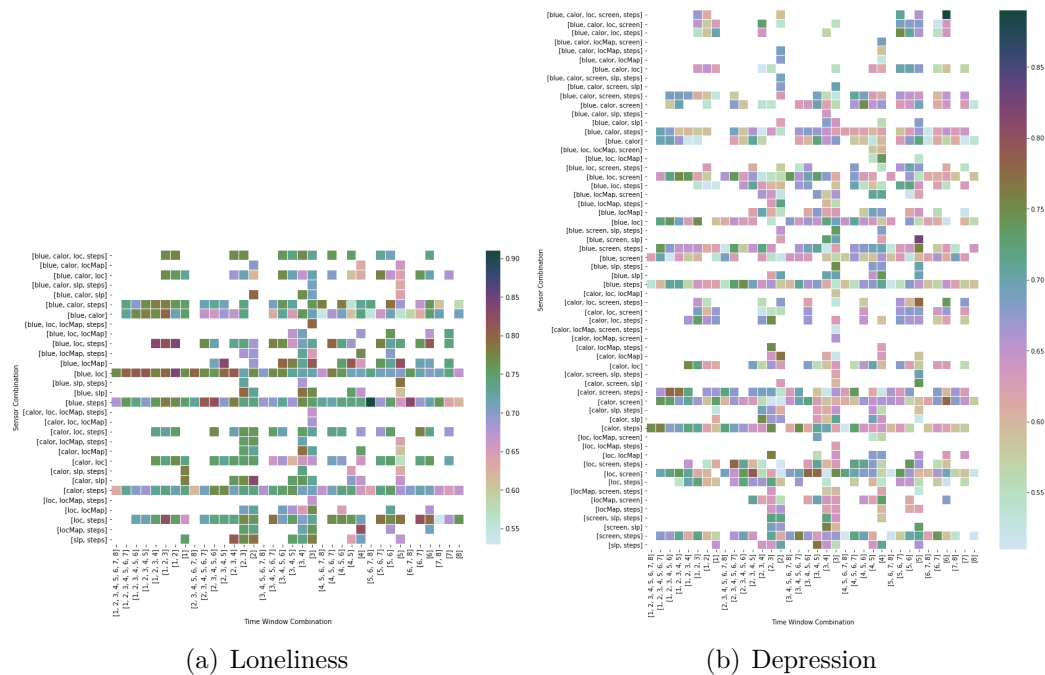


Figure 4.5: The heatmap displays the largest F1 score in the loneliness prediction model trained by a combination of different multiple sensor features and time windows.

the multiple sensor modeling, I observe that the maximum F1 scores of predicting loneliness and depression are 0.91 and 0.89, which are obtained from the combination of Logistic Regression and Lasso. Overall, for the majority of approaches, the combination of Gradient Boosting and Information Gain provides the best performance. This combination should be further evaluated with other similar datasets to replicate and confirm their superior performance over other algorithm combinations.

	Single Sensor						Multiple sensors					
	Loneliness mean(max)			Depression mean(max)			Loneliness mean(max)			Depression mean(max)		
	GB	LR	RF	GB	LR	RF	GB	LR	RF	GB	LR	RF
IG	0.69 (0.76)	0.69 (0.76)	0.66 (0.72)	0.58 (0.70)	0.60 (0.61)	0.56 (0.63)	0.73 (0.83)	0.72 (0.78)	0.69 (0.81)	0.63 (0.83)	0.60 (0.66)	0.63 (0.76)
Lasso	0.68 (0.72)	0.70 (0.76)	0.74 (0.74)	0.57 (0.68)	0.57 (0.64)	0.55 (0.59)	0.72 (0.78)	0.75 (0.91)	0.59 (0.66)	0.67 (0.89)	0.54 (0.54)	0.54 (0.54)
RLL		0.70 (0.76)	0.68 (0.73)	0.58 (0.65)	0.56 (0.65)	0.57 (0.60)	0.75 (0.81)	0.73 (0.82)	0.76 (0.84)	0.65 (0.78)	0.65 (0.79)	0.65 (0.79)

Table 4.5: Summary of the mean and maximal values of F1 score for each combination of feature selection and machine learning methods shown in the heatmaps 4.4, 4.5. The bold values are either the biggest mean value of F1 scores, or the biggest maximal values of F1 scores

	TW1	TW2	TW3	TW4	TW5	TW6	TW7	TW8	HDominant
Audio	Amp SE	Mesor SE	Amp SE	IPR	Magnitude	Amp SE	Bathypase	P	Amp SE
Battery	IPR	PR	Mesor SE	Mesor SE	Orthophase	Magnitude	Orthophase	Bathypase	Mesor SE
Bluetooth	Magnitude	Bathypase	Amp	P	IPR	Orthophase	Mesor SE	Orthophase	Orthophase
Call	IPR	PHI	IPR	IPR	Amp SE	Bathypase	Orthophase	Magnitude	IPR
Calorie	Mesor	Magnitude	Magnitude	Bathypase	Orthophase	Orthophase	IPR	Magnitude	Magnitude
Location	PHI SE	Magnitude	Mesor	PR	IPR	Mesor	Amp SE	IPR	Mesor
Location Map	Orthophase	Magnitude	Mesor	Orthophase	PHI	Bathypase	Orthophase	Bathypase	Orthophase
Messages	Orthophase	Magnitude	LCM	PR	Mesor SE	Bathypase	PHI SE	Magnitude	Magnitude
Screen	Amp	P	Orthophase	Orthophase	PR	Orthophase	IP	Amp SE	Orthophase
Sleep	Bathypase	PHI SE	Mesor	Orthophase	PHI SE	IP	Amp SE	Bathypase	Bathypase
Steps	P	Orthophase	Magnitude	Bathypase	PR	IPR	IPR	Magnitude	Magnitude
Wifi	Amp	Mesor SE	Mesor	Orthophase	Magnitude	IPR	IP	Amp SE	Magnitude
VDominant	Amp	Magnitude	Mesor	Orthophase	Orthophase	Bathypase	Orthophase	Magnitude	Orthophase

Table 4.6: The most frequently selected rhythmic features by Information Gain during depression prediction.

Dominant rhythm parameters that predict mental health I count the frequency of rhythmic features selected by machine learning models to measure the contribution of each rhythm parameter in predicting mental health. Orthophase and Magnitude appeared on top of the list as the most frequently selected parameters. Although I used three feature selection methods in the evaluation, I observed that the Information Gain method provided more reliable and complete list of features during the training. Table 4.6 shows the rhythm features that are selected most frequently by Information Gain during depression prediction for each sensor feature in each time window. The vertical dominant feature (VDominant) is the most commonly selected feature for most of the sensors at a given time window, and the horizontal dominant feature (HDominant) is the most commonly selected feature in most time windows for a given sensor. The overall dominant feature (the feature at the bottom right corner in bold font) is the most commonly selected feature for all sensors and time windows. If two features are the most commonly selected features for the same number of sensors/time windows, I break the tie by taking the feature with a higher frequency. Overall, Orthophase is selected most frequently for all sensors and time windows. Magnitude comes the second. Given that Phase and Magnitude reflect duration and intensity of biobehavioral features, frequent selection of these parameters suggests an

important relationship with mental health status.

In addition to main rhythmic features, i.e., Mesor, Amplitude/Magnitude, and Ortho/Bathypphase, I observe frequent selection of features related to the fit of Cosinor models including the significance level of the fit (P), Standard Errors (SE) and Percent Rhythm (PR and IPR), i.e. the proportion of the overall variance accounted for by the fitted model. Higher levels of these parameters reflect higher variation in data. Therefore, frequent selection of these parameters indicates the power of regularity/irregularity of biobehavioral rhythms in predicting mental health status.

Comparison with Models Built without Rhythm Parameters

To better understand the capability of the framework in utilizing rhythmic features to predict an outcome, I compare the prediction performance of the models with rhythm modeling against the models without rhythm modeling. Specifically, I select the best performing sensor feature in each time window, run exactly the same machine learning pipeline on the raw feature data without rhythm modeling, and compute the F1 score. Table 4.7 shows that the pipeline with rhythm modeling outperforms the one without by a large margin on most of the features. This observation is consistent with both loneliness and depression predictions.

Time Window	Feature	Rhythm-F1	Raw-F1
1	shortest period spent at Halls	0.66	0.54
2	longest awake period length	0.64	0.49
3	number of awakes	0.63	0.47
4	maximum calories increase between 5-min periods	0.66	0.60
5	shortest asleep period length	0.70	0.69
6	total distance traveled	0.65	0.50
7	maximum calories decrease between 5-min periods	0.67	0.59
8	minutes spent at Halls	0.65	0.62

Time Window	Feature	Rhythm-F1	Raw-F1
1	shortest period spent at Halls	0.69	0.55
2	longest awake period length	0.67	0.47
3	total asleep time	0.67	0.49
4	number of awakes	0.62	0.56
5	percentage of time spent moving	0.72	0.52
6	longest period spent at athletic areas	0.68	0.43
7	total change of calories	0.68	0.53
8	variance of moving speed	0.67	0.48

Table 4.7: F1 of machine learning models with rhythm modeling (rhythm) and without rhythm modeling (raw features). Left: Loneliness; Right: Depression.

4.1.2 Case2: Predicting Readiness

A second dataset was selected to assess the framework's adaptability in modeling various data types and employing different analysis approaches. The sensors, participants, and ground truth for this dataset differ from the previous one. In this case, I used data from 11 undergraduate and graduate students who continuously wore Oura ring, a commercially available smart and convenient health tracker for several months. As shown in the last plot of Figure 4.7, the length of data collection varies per participant and ranges from 31 to 323 days. The long-term data makes it possible to detect and observe rhythms with larger cyclic periods than a day, e.g., weeks or months. As such, I design the analysis to answer the following:

1. Are there common cycles in participants' data per sensor and across sensors, and can I identify similarities and differences in cyclic periods among participants despite differences in the length of their data?
2. How accurately can individual rhythm models per sensor feature and per participant predict average readiness?

Physiological Data Processing

OURA collects sleep, heart rate, skin temperature, calories, steps, and activity. Sleep, heart rate, and skin temperature samples are collected every five minutes during night hours; and activity, calories, and steps are sampled every 5 minutes during the day. The data is summarized and stored on the OURA cloud platform. As the goal is to detect cycles with multiple-day lengths, I aggregate the features into daily intervals (as opposed to the previous case that used hours). In total, I use 31 features such

as total duration of sleep, lowest/average heart rate, average metabolism level, total amount of calories burned, and total number of steps during the day. To be able to detect the longest periods in participants' data, I refrain from segmenting data into common time windows and use the entire time series data for the analysis. The convenience of wearing the ring and its long battery life lead to good quality data with low missing rates (Max 15.6% in the data). I use the moving average method to handle the missing values.

Readiness Score as Ground Truth

Besides the physiological features, Oura provides a readiness score displayed in the Oura ring app, i.e., an evaluation of the body's overall recovery rate after waking up in the morning. The readiness score ranges from 0 to 100 with scores over 85 indicating high readiness for challenging tasks and scores below 70 indicating poor body state and need for recovery. In the dataset, participants' readiness scores range from 24 to 99 with an average score of 74, and a standard deviation of 11.4. Figure 4.7 and 4.6 shows the distribution and variation of daily readiness score for each participant. I calculate the average daily readiness score for each participant and use it as ground truth to explore how well I can use the rhythms to predict the readiness score.

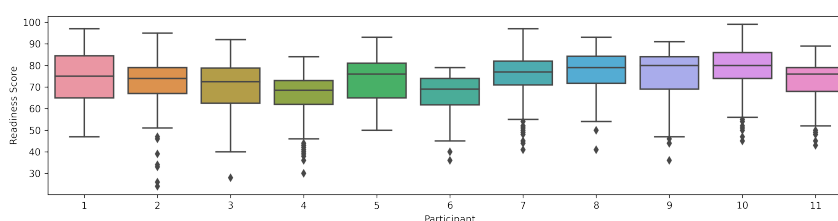


Figure 4.6: The 1 to 11 boxplots display the minimum, median, maximum, and quartile of the daily readiness scores for each participant. Most daily readiness scores are clustered in the range from 70 to 85.

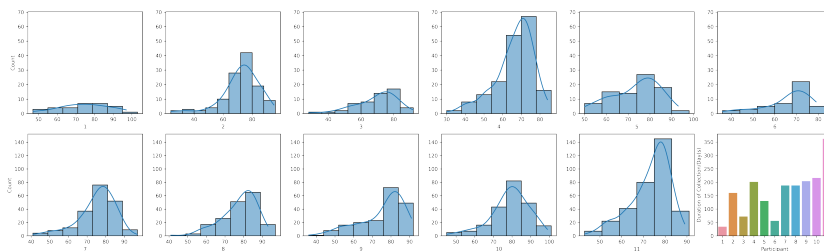


Figure 4.7: The histograms from 1 to 11 display the distribution of the daily readiness scores for each participant, and the last bar plot shows the duration of each participant’s data collection.

Detection of Cycles in OURA-Ring Data

the first analysis questions relate to detecting common cycles in the participants’ data and physiological sensors. the results show weekly and biweekly periods were observed most frequently. Similar to case 1, I applied Fourier Periodogram on the time series data of each sensor feature per participant to detect significant periods. In Table 4.8 and 4.9, I list the most frequently detected periods of sensor features and summarize them by sensor type and participants. The number 7 and its multiple 14 as well as its close preceding and following numbers of 6 and 8 appear most frequently in both tables suggesting near-weekly bibehavioral patterns. In particular, periods of Activity, Sleep, and Heart rate project near-weekly cycles across all participants. For example, Activity cycles of 6, 7, and 8 days are observed in 45%, 55%, and 36% of participants respectively. These cycles are also observed in sensor data of seven participants (63%). Calorie and Steps share periods of 2, 10, and 11 days with similar percentages. Although the percentages of participants with these cycles are low likely due to different movement patterns among participants, the common periods of these two sensors may be indicative of exercise cycles in those participants.

Sensor	Detected Period (% of Participants)
Activity	7 (55), 2 (45), 6 (45), 8 (36), 4 (36)
Calorie	2 (18), 11 (18), 10 (18), 4 (9), 81 (9), 20 (9)
Heart Rate	7 (36), 27 (27), 8 (27), 14 (18), 18 (18)
Sleep	8 (55), 3 (55), 7 (45), 6 (45), 11 (36)
Steps	11 (27), 10 (27), 2 (18), 54 (18), 7 (18)
Skin Temperature	12 (36), 14 (36), 15 (27), 27 (27), 34 (18)

Table 4.8: Dominant frequent periods for each sensor. The percentage in parenthesis is the percentage of participants with the significant period.

Participant	Detected Period (% of Sensor Features)
1	7 (29), 34 (26), 2 (23), 3 (16), 39 (10)
2	80 (42), 81 (39), 40 (35), 11 (29), 32 (26)
3	77 (32), 10 (29), 24 (23), 7 (23), 26 (16)
4	7 (52), 202 (39), 101 (35), 67 (19), 201 (16)
5	66 (39), 65 (35), 130 (26), 8 (26), 26 (23)
6	6 (35), 56 (29), 14 (23), 28 (13), 19 (10)
7	31 (26), 11 (23), 190 (23), 95 (23), 38 (19)
8	94 (42), 188 (29), 63 (29), 7 (23), 189 (23)
9	68 (45), 102 (35), 29 (29), 204 (26), 41 (16)
10	54 (45), 108 (39), 43 (35), 27 (23), 217 (32)
11	126 (35), 42 (26), 28 (23), 5 (16), 7 (16)

Table 4.9: Most frequent periods of all sensor features for each participant. The percentage in parenthesis is the percentage of sensor features with that period.

Prediction of Readiness with Rhythmic Features

For each participant, I use the three most significant periods identified by the Periodogram as input to the Cosinor method to build rhythm models per sensor feature. The rhythmic features are then entered into the machine learning process to predict average readiness per participant. Since the readiness score is a continuous variable, I build regression models to make predictions. The choice of machine learning algorithms includes Random Forest and Gradient Boosting with Information Gain and Lasso as feature selection methods. Similar to case 1 in mental health, I build models with single and multiple sensor combinations in a leave-one-participant-out cross validation, but, instead of accuracy, I use the Root Mean Square Error (RMSE) as the performance measure.

Table 4.10 lists the best RMSE achieved by single sensor models along with the most frequently selected features. Among single sensor models, the model built with

the rhythmic feature of sleep data with an RMSE of 4.08 is a stronger predictor of readiness than others. In comparison, the combination of sleep, calories, and steps obtain an RMSE of 3.54, the lowest RMSE among all multiple sensor models, as shown in Table 4.11. This combination considers both the activity of the human body during the day (calories) and the sleep quality at night (sleep). These observations are expected and confirm the impact of both sleep and physical activity on the body’s daily functioning. Interestingly but not surprisingly, the frequently selected features across all sensors are standard errors of the rhythm parameters (i.e., PHI SE, MESOR SE, and Amp SE) as well as percent rhythm (PR), all of which are indicative of variation in the actual data. MESOR SE is the most dominant feature among both single and multiple sensor models. These results suggest that the level of variability and potentially irregularity in biobehavior may be most predictive of fluctuations in readiness.

Table 4.10 and 4.11 also summarize the RMSE for models using each combination of feature selection and machine learning methods. The Gradient Boosting model with Lasso regression achieves the best performance for both single sensor and multiple sensor modeling, with an RMSE of 3.54. Using the same prediction model, the Information Gain performs better in single sensor modeling, and the results are reversed in multiple sensor modeling.

Sensor	Activity		Calorie		HR		Sleep		Step		Skin Temperature	
Feature Selection	IG	Lasso	IG	Lasso	IG	Lasso	IG	Lasso	IG	Lasso	IG	Lasso
RMSE (GB)	5.04	8.42	4.79	5.18	4.54	5.50	4.08	5.54	4.71	6.77	5.34	6.77
RMSE (RF)	5.25	8.52	4.38	4.51	4.65	6.20	4.20	5.68	4.81	7.30	5.48	7.30
Frequent Rhythmic Features	PR	PHI	Mesor SE, Amp SE	PHI	PR	PHI, PHI SE, P	PR	P	Mesor SE, Amp SE	Mesor, P	Mesor SE, Amp SE, P	PHI

Table 4.10: Lowest RMSE of single sensor features and frequent rhythmic features selected by IG and Lasso

Feature Selection	IG	Lasso
Sensor	sleep, calorie, step	sleep, calorie, step
RMSE (GB)	3.73	3.54
RMSE (RF)	3.80	3.68
Frequent Rhythmic Features	MESOR SE	MESOR

Table 4.11: RMSE of multiple sensor models and frequent rhythmic features selected by those models.

4.2 Identifying Relationship between Rhythms and Productivity

In this section, the correlation method detailed in Section 3.4 demonstrates its effectiveness in identifying associations between sensor-specific-rhythm stability and overall-rhythm stability with high and low productivity. The dataset utilized for this analysis is a dataset of smartphone and Fitbit logs collected from 188 students at an American university over a semester was used, with 166 students providing subjective evaluations of their daily productivity. Data collection involved using the AWARE app [108] and Fitbit devices to gather information on audio, Bluetooth, Wi-Fi, location, phone usage, calls, calories, sleep, and steps. All participants utilized their smartphones, while the research team provided a Fitbit Flex2 to gather data.

Students' productivity assessments were collected via an evening survey during week one, six (mid-semester), and fifteen (last week of classes) of the semester to avoid overburdening participants. The assessment question included a single question: "How productive did you feel today?". The possible responses ranged from 0 (not productive at all) to 4 (extremely productive). The mean and standard deviation of self-evaluated productivity scores were consistent for different gender and major groups with no significant difference: female (1.65, 0.92), male (1.80, 0.97), engineering (1.71, 0.96), business (1.70, 0.99), science (1.69, 0.94), art (1.76, 0.95), humanities (1.68, 0.97), undecided (1.67, 0.87)). The number of responses varied across the three weeks as

some participants did not complete the surveys. In total, the dataset contained 488 observations (participant-week pairs).

I used productivity scores (0-4) to categorize participants into high and low productivity groups. These categories were used as ground truth labels in the machine learning classification. To identify the cut-off threshold, I calculated the mean and median of the daily productivity scores for all participants across all three weeks. The mean of 1.89 and a median of 2 indicated a normal distribution across scores (verified by the Shapiro-Wilk test, $p = 0.12$). Therefore, I used 2 as the threshold for categorizing productivity, with scores less than 2 indicating low productivity and scores equal to or above 2 indicating high productivity. The variance of productivity scores within each week was less than 1, which indicated that participants' productivity assessments were relatively stable each week. We, therefore, averaged the productivity scores of all days in each week (including both weekday and weekend) as the weekly productivity score with the same threshold to categorize each participant's week average into high or low productivity. I further categorized participants into high and low productivity groups, where students with at least two weeks (out of three) of high productivity rates were categorized as highly productive, and the rest were placed into the low productivity group.

I then used Cosinor to build personal cyclic models per student per sensor stream in weeks 1, 6, 15, and the weeks adjacent to them (e.g., for week 6, I use sensor data from week 5, 6, and 7 to build Cosinor models). I then used rhythm parameters generated by those models in the correlation analysis and the machine learning pipeline. I assumed all participants had normal daily rhythms and used the input periods of 8h, 12h, 24h in the Cosinor. The 8, 12, and 24 hours reflect nocturnal, diurnal, and circadian duration, respectively.

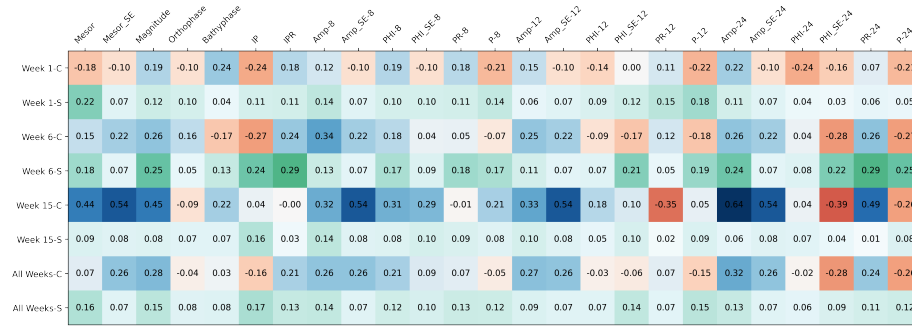


Figure 4.8: The heatmap displays correlations between rhythm parameters and productivity by week. The blue (positive correlation) and red (negative correlation) cells show the correlation average (C-RF) by week (Week-C), and the green cells show the aggregated significance score (S-RF) by week (Week-S).

While correlations between rhythm parameters and productivity scores were moderate across all behavioral sensor features and all three weeks (Figure 4.8), I observed more pronounced relationships between parameters related to regularity in rhythm models, including standard errors (SE), i.e., deviation of the fitted model parameter from the actual values, Percent Rhythms (PR and IPR) or proportion of variation accounted for by the fitted model, and the significance of the fit (P and IP). In addition, the aggregated negative correlation (indicated by the red line) in the majority of these parameters across all three weeks indicates lower rhythm irregularity in highly productive students.

The rhythm parameters for location features appeared to be dominant in both aggregated correlation coefficients and significance scores, followed by activity and sleep features (Figure 4.9). In the following, I discuss the observations in detail.

4.2.1 Correlation Aggregation of Rhythm Parameters

The blue and red cells in Figure 4.8 show the correlation aggregated by week for each rhythm parameter as calculated using equations 3.9, 3.10, 3.12 and 3.14 in Ap-

pendix ??). Recall that these formulas aggregate correlation across all sensor features for each rhythm parameter to measure the strength of the correlation between productivity and the rhythm parameter. Blue cells indicate positive correlation while red cells indicate negative correlation.

The green cells in Figure 4.8 show the significance score by week for each rhythm parameter as computed by equations 3.11 and 3.13 in Appendix ??. These formulas calculate correlation significance across all sensor features for each rhythm parameter to measure the significance of the correlation between productivity and the rhythm parameter. The higher the significance score, the more significant the relationship is.

Week 1 In week 1, the majority of parameters that measure the irregularity of the rhythm models correlate negatively with productivity indicating more stable rhythms in the high productivity group. For example, stronger correlations were observed between productivity and the model fit for the fundamental period (IP) ($C = -0.24$), the 24-hour period (P-24) ($C = -0.21$), the 12-hour period (P-12) ($C = -0.22$), the 8-hour period (P-8) ($C = -0.21$), the fundamental percent rhythm (IPR) ($C = 0.18$), and standard error (SE) of phase fit for the 24-hour period (PHI_SE-24) ($C = -0.16$).

The relationship between regularity in rhythms and productivity is further reinforced by the negative aggregated correlation coefficients for P-24, P-12, P-8, IP, and SE. Specifically, their low values indicate that Cosinor was able to create close fits to the actual data which means more regularity in the actual data corresponds to high productivity. This further demonstrates a lower rhythm variation in highly productive students.

The relationship between lower rhythm variability and higher productivity is also observed in the correlation of Mesor_SE, Amp_SE-8, Amp_SE-12, Amp_SE-24,

and PHI_SE-24. The values have a relatively high aggregated significance score compared to other parameters. This means the standard error (SE) has a more significant relationship with productivity. Given the SE is also a metric reflecting irregularity of rhythm models, its negative correlation indicates less irregularity of the rhythm models in high productivity.

The percent rhythm parameter also demonstrated a relationship between low rhythm variability and high productivity. A higher percent rhythm represents low variability in the actual data. Specifically, the percent rhythm of the fundamental, 24-hour, 12-hour, and 8-hour period all have high positive aggregated correlation coefficients with productivity, indicating lower variability in diurnal activities for the highly productive students.

Week 6 Week 6 (midterm) projected a relatively different pattern. For example, I found positive correlations between productivity and Mesor_SE, Amp_SE-8, Amp_SE-12, and Amp_SE-24. Because amplitude and Mesor are indicative of the intensity and volume of activities, I see highly productive students performed more intense activity during week 6.

I also found amplitude and Mesor have higher standard errors (SE) in the fitted models. This implies higher variability in the intensity of regular patterns during this week. This can be expected due to midterm pressure.

Despite this increased variability of intensity of regular activities, as demonstrated by the positive aggregated correlations of IPR ($C = 0.24$) and PR-24 ($C = 0.26$) with productivity, I see less irregularity in activity patterns during this week for the highly productive students.

Finally, like week 1, I see positive correlations between percent rhythm (PR) and productivity. However, the correlation became more stable in week 6 compared to week 1 with larger aggregated significance scores.

Week 15 Week 15 (the week before finals) showed the the strongest correlations. For example, parameters that reflect irregularity in rhythms such as Standard Errors (SE) (e.g., Mesor_SE, Amp_SE, PHI_SE) show high (mostly positive) correlations with productivity. Parameters characterizing the fitted cyclic model such as Mesor, Phase, and Amplitude also show high (mostly positive) correlations with productivity indicating higher intensity and duration of behavioral activities during this week.

The value of some correlations, however, decreased from weeks 1 and 6 to week 15. For example, the correlation between Percent Rhythms (e.g., IPR PR_8, and PR_12) and productivity. Given the increased workload activities close to final exams, the observed irregularity and divergence from the routine patterns are expected.

Despite the decline in the value of some correlations, observations across all three weeks still suggest an overall lower irregularity in rhythms among the high productivity group. For example, there is a consistent negative correlation of the regularity indicators such as P-24, P-12, P-8, PHI-SE-24, PHI-SE-12, PHI-SE-8, and IP. Moreover, parameters representing the phase's characteristics in rhythms including Orthophase, Bathyphase, PHI-24, PHI-12, and PHI-8 exhibit relatively high aggregated significance scores in all three weeks. This means more regularity in phase is more significantly correlated with high productivity. Thus, while further explorations are needed, these observations indicate the importance of rhythm stability in students' productivity.

4.2.2 Correlation Aggregation of Sensor Features

Figure 4.9 shows the aggregated correlation and significance scores by week for the top ten sensor features calculated through equations 3.3, 3.4, 3.5, 3.6, 3.8 and 3.7 in Appendix ???. These formulas calculate the aggregated correlation coefficients and significance scores across all rhythm parameters for each sensor feature to measure the strength of the correlation between productivity and the behavioral sensor features. Features with higher significance scores have a more significant correlation with productivity. Overall, location features had a stronger aggregated correlation and significance. The rhythm model for each sensor feature was not consistently associated with productivity in all three weeks.

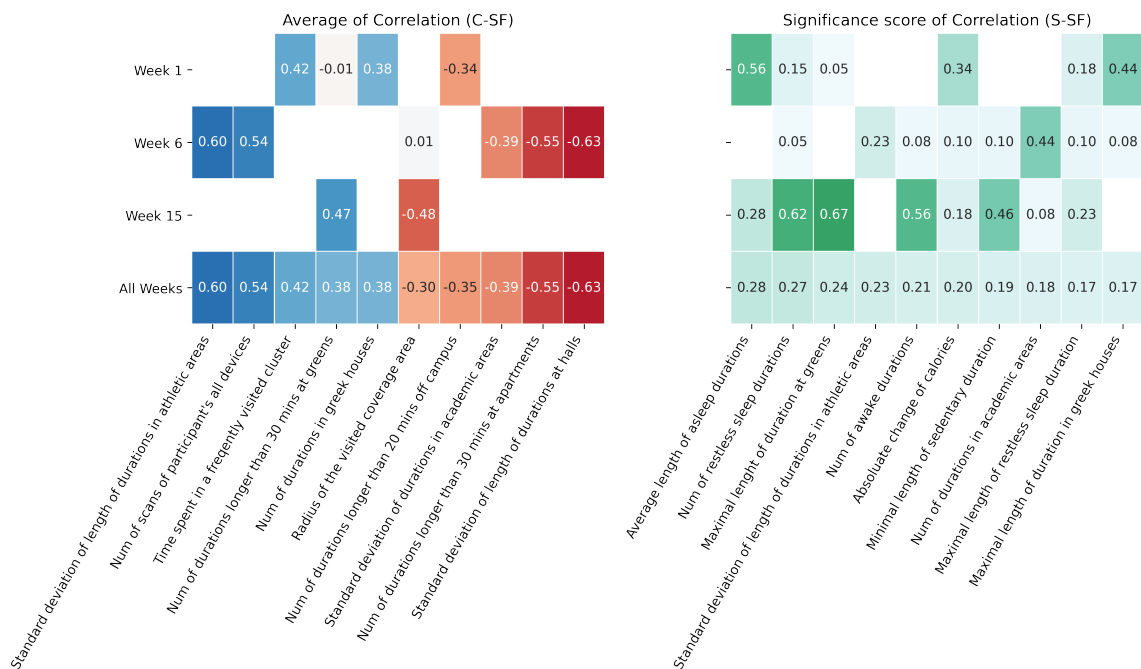


Figure 4.9: The heatmap displays the correlation between sensor features and productivity by week. The left side shows the ten sensor features with the highest aggregated correlation over all three weeks, and the right side shows the ten sensor features with the highest aggregated significance score over all three weeks.

Week 1 In week 1, rhythm parameters for both the time spent in frequently visited places and the frequency of visits in fraternity/sorority houses (places for socializing) showed the highest average positive correlations with productivity. A negative correlation between productivity and off-campus duration was also observed in the rhythm models. Finally, I found patterns of asleep and burned calories to have high significance scores.

Week 6 In week 6, the variance of the length or number of stays in academic areas, halls, and apartments showed high negative aggregated correlations with productivity (the left side of Figure 4.9), indicating that highly productive students had a stable living and studying environment at home and school. Conversely, the standard deviation of duration in athletic areas were positively correlated with productivity. This indicates higher variability in exercise associated with high productivity. A similar conclusion can be drawn with the data from the aggregated significance score data (the right side of Figure 4.9).

Week 15 In week 15, I observed the highest aggregated significance scores for rhythms of restless sleep duration, awake sleep duration, time spent at greens, and sedentary duration. On the left side of Figure 4.9 I see the time spent at greens was positively correlated with productivity, whereas the radius of the visited areas was negatively correlated with productivity. This finding suggests that high-efficiency students reduced their range of activities and spent time outdoors more frequently in week 15.

4.3 Measuring Variability in Biobehavioral Rhythms

In this section, I explore various applications of measuring variability in biobehavioral rhythms. First, I apply change point detection methods to physiological signals in order to identify when disruptions occur within individual physiological cycles. Next, I use the COSANOVA method to quantify rhythm variability between different readiness or well-being groups, as well as across time. This investigation demonstrates the potential applications of assessing variability in biobehavioral rhythms, offering valuable insights into the complexities of human behavior and its impact on health and well-being.

4.3.1 Identifying Rhythm Disruption in Physiological Cycles

As shown in Appendix A, the FFT, chi-square periodogram, and CyHMMs have been validated in dataset 1 (synthetic data), and they are all able to identify a 24-hour rhythm period for heart rate and skin temperature in datasets 2 and 3. I continue the experiment on all three datasets to further detect the inner cycle and change points with the AutoNOM method (mentioned in Section 3.6.1) within each period.

Dataset 1 is generated using the combination of constant frequencies and noise, and the AutoNOM identify changing points by perceiving changes in frequency, so it is reasonable that no change points have been detected in Fig 4.10. From another perspective, the AutoNOM has a strong anti-noise ability when searching for change points.

To demonstrate the performance of AutoNOM in detecting changing points in dataset 2, I choose one day from all 70 days in the dataset and use the AutoNOM to fit the

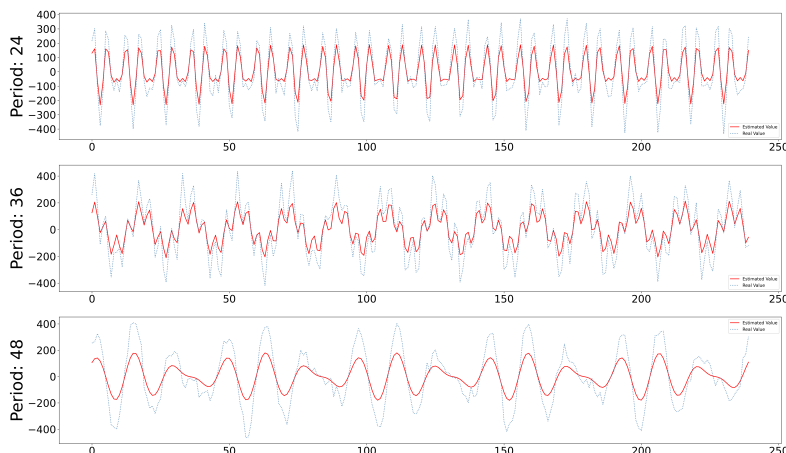


Figure 4.10: The analysis of dataset 1 using the AutoNOM. The blue lines are the real data, and the red lines depict the estimated values from the AutoNOM. No changing points have been detected by the AutoNOM, which is expected as there are no abrupt frequency changes in the synthetic dataset.

one day's data. I choose k equal to 4 as the optimal value based on an empirical observation: as shown in Table 4.12, when the value of k changes from 3 to 4, the MAPE value drops significantly, and when the value of k changes from 4 to 5, the change is minor. Figure 4.11 demonstrates similar distributions of the heart rate and the skin temperature change points where the two most frequent change points are located at around the 400th minute (6:40 am) and 1200th minute (8:00 pm) for both distributions. Illustrated in Figure 4.12, the skin temperature remains high at the start and end of each day, and it has a trough in the middle. As for the heart rate, it keeps low at the beginning of each day and goes through two combinations of up and down in turn and returns to a lower value at the end of the day. Between the two most frequent change points, the amplitudes and frequencies of the oscillation increases. Based on previous research, the body temperature of the circadian rhythm is under the control of the suprachiasmatic nucleus (SCN), which receives the input from photosensitive cells and synchronize body temperature and day and night alternation [9]. The temperature will reach its peak in the late afternoon and drop to its

trough at the end of sleep [43]. In Figure 4.12, an abrupt decrease in skin temperature at the second changing point (about 7:00 am) and a stable increase between the second (about 7:00 am) and third changing point (about 5:00 pm) can be observed, which is consistent with previous studies. However, the difference is that the peak of skin temperature occurs in the late evening. Compared with body temperature, heart rate is less susceptible to the influence of the external environment [32, 20]. The heart rate is more dependent on physical activity, and the heart rate during sleep is lower than during the wake time [135]. This view can explain why the heart rate in Figure 4.12 is lower on both sides of the day. The increase in heart rate in the two segments may be related to physical activity.

	k=3	k=4	k=5
Heart Rate	4.54	3.29	3.14
Skin Temp	3.10	1.92	1.73

Table 4.12: The MAPE values of the AutoNOM method with different maximal number of change points k .

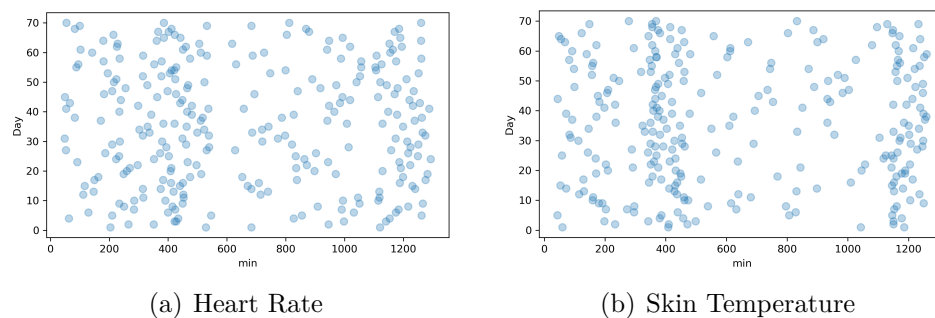


Figure 4.11: The distribution of change points in heart rate per day for 70 days. The x-axis represents the time of day in minutes. The change points are concentrated in the 400th and 1200th minutes of each day. Compared with the heart rate, the temperature has a more obvious tendency of the concentrated distribution. This change in heart rate and skin temperature is similar to the pattern of rest activity alternating between daytime activity and night rest.

I repeated the same process for dataset 3 and did the same analysis process on one

random day in the data. The grey vertical lines in Figure 4.13 illustrates the changing points detected by the AutoNOM for heart rate, skin temperature, EDA, and BVP on the chosen day. The two common changing points for all the signals occur between the 400th minute (6:40 am) and 600th minute (10:00 am) and at around 1200th minute (8:00 pm), which are similar to the observation from dataset 2. In each segmentation between two changing points, the estimated curve output from the AutoNOM could fit the raw data well except for BVP. BVP reflects the relative change in blood volume caused by the heart contracting, so there will be many instantaneous and massive changes in BVP signals, which cause the AutoNOM not to work well. The daily trend of heart rate and skin temperature shown in Figure 4.12 and 4.13 are different, which could be caused by individual differences (e.g., lifestyle, daily schedule, and personality traits), climate, and even accuracy of wearable devices. I observe a seasonal effect in the skin temperature between the two datasets. The skin temperature in Figure 4.13 shows an apparent decrease between the second and third changing points, whereas the skin temperature in Figure 4.12 shows an increase at the same time. Checking the timing of data collection for these datasets, I found that dataset 2 was collected during spring, whereas dataset 3 was collected during the early winter. One interesting observation from dataset 3 is that there is one peak for heart rate during the first and second changing point, which is inconsistent with what I mentioned above. Peters et al. have found that the accelerated heart rate during sleep may be caused by uncomfortable sleeping posture [49]. When in an uncomfortable posture, the volume of intake oxygen will decrease, so the heart will increase the beat rate to demand oxygen supply, which is similar to what happens during strenuous exercise [146].

Empatica E4 used in the dataset 3 is a medical-grade device that can collect accurate

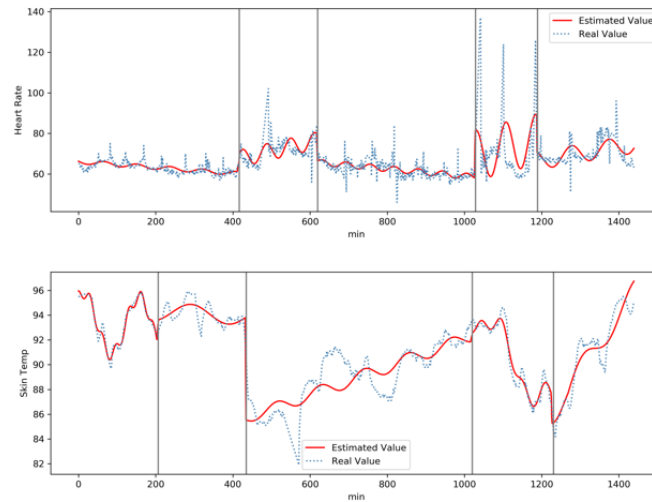


Figure 4.12: The analysis of heart rate and skin temperature in dataset 2 for one day using the AutoNOM. The blue lines are the real data, and the red lines depict the estimated values. The gray vertical lines show the estimated location of the change points. Three changing points occur at around 400th, 1000th, and 1200th in both heart rate and skin temperature. The heart rate increases in the second and fourth segments, while the remaining three segments keep low. The skin temperature maintains a high value at the beginning and end of the day, and there are a significant decrease and rebound in the middle.

physiological data, but E4 can only work up to 40 hours and needs to be charged again. E4 cannot work when charging, and this process introduces missing values, which will cause unnecessary changing points. For example, the time points when E4 stops and starts working will be recognized as changing points. Due to the above uncontrollable factor, I do not provide a figure similar to Figure 4.11 for dataset 3.

4.3.2 Quantifying Rhythm Variability of Readiness and Well-being Outcomes

To assess the effectiveness of the COSANOVA method introduced in Section 3.6.2, I validate the approach using two real-world datasets, and the variability of rhythms is then explored in two extensive, multimodal mobile sensing datasets. I choose these

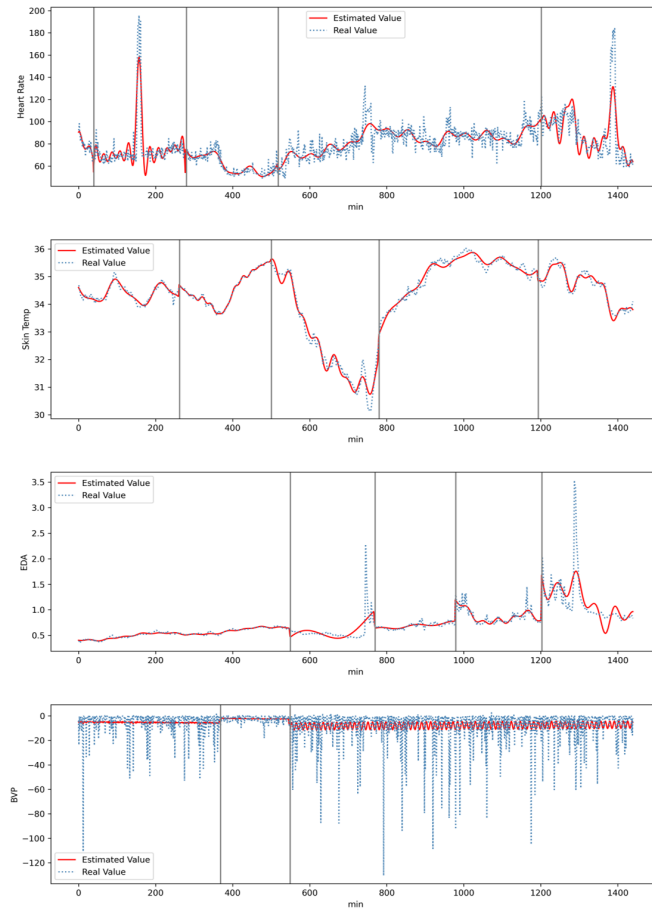


Figure 4.13: The analysis of heart rate, skin temperature, EDA, and BVP in dataset 3 for one day using the AutoNOM. The blue lines are the real data, and the red lines depict the estimated values from the AutoNOM. The gray vertical lines show the estimated location of the change points. The changing point between 400th and 600th minute is pretty close in the five signals. Heart rate, skin temperature, and EDA all have a changing point approximately at the 1200th minute. The heart rate fluctuates greatly at both ends, and the peak occurs at the beginning of the day. Similar to Figure 4.12, the trough of skin temperature appears in the middle. Except at the end of the day, EDA has little fluctuation. BVP declines in the second segment and has many instantaneous changes.

datasets because they were collected continuously by mobile phones and wearable devices for a long term, and they have objective or subjective reports of health or well-being conditions. Through exploratory application of COSANOVA on the two human behavior dataset, I plan to answer the following questions:

- Can different rhythm stability scores be observed on different populations or different time windows?
- Is there any associations between rhythm stability scores and mental health and readiness scores.
- Which sensor feature and rhythm parameter are in the dominant position when measuring stability scores?

Case1: Quantifying Rhythm Variability on Real-world Readiness dataset

In this case, the Oura ring dataset, previously utilized in Section 4.1.2, is employed. The Oura ring measures daily readiness scores based on physical activity and physiological metrics. Collected data includes sleep, heart rate, skin temperature, calories, steps, and activity. Features are aggregated into daily intervals and analyzed without segmenting into common time windows, aiming to detect the longest periods in participants' data. Since there are no ground truth of periodic patterns for the sensing signals in the readiness and well-being datasets, it's impossible to evaluate the performance of COSANOVA on the two real-world datasets. Instead, I directly explore the feasibility of application of COSANOVA on human behavioral data and summarize how human behaviors varies between different health groups (VANOVA) and over time (HANOVA). Figure 4.15 represents the results of COSANOVA for the top 10 features in readiness dataset by the variability when calculating VANOVA and HANOVA respectively. Features with a low COSANOVA value have more variability in specific periodic components. Overall, sensor features related to sleep had stronger variability between the two readiness populations and over time. Mesor is more likely

to be significant different compare to the other two rhythm parameters.

	HANOVA			VANOVA		
	Mesor	Amp	Phi	Mesor	Amp	Phi
sleep_hr_lowest	4.4e-02	4.3e-01	1.9e-02	8.8e-04	1.3e-01	3.6e-02
sleep_score_latency	4.4e-02	1.0e-01	1.9e-02	8.8e-04	1.3e-01	3.6e-02
sleep_temperature_deviation	1.2e-07	9.1e-02	1.7e-01	8.8e-04	9.2e-02	1.2e-01
sleep_score	1.2e-36	1.5e-01	1.4e-02	1.8e-03	4.6e-01	2.1e-01
sleep_score_alignment	9.6e-35	1.9e-01	1.4e-02	4.0e-04	4.9e-01	2.3e-01
activity_target_miles	3.5e-75	3.3e-02	4.2e-01	8.6e-03	3.9e-01	3.3e-01
sleep_efficiency	1.1e-30	2.2e-01	2.7e-03	1.0e-02	5.1e-01	2.1e-01
sleep_score_efficiency	1.8e-40	1.7e-01	1.0e-02	2.1e-01	3.3e-01	2.2e-01
sleep_onset_latency	1.8e-43	8.2e-02	7.8e-03	5.1e-02	4.2e-01	3.0e-01
activity_to_target_km	1.9e-79	3.3e-02	4.6e-01	2.2e-03	4.0e-01	3.8e-01
sleep_temperature_trend_deviation	1.1e-44	3.7e-02	1.1e-02	1.3e-01	5.2e-01	3.3e-01
sleep_score_deep	1.1e-44	2.4e-02	6.4e-02	1.3e-01	5.5e-01	3.7e-01
sleep_score_rem	4.4e-52	1.7e-02	8.0e-02	1.2e-01	4.5e-01	3.4e-01
sleep_duration	6.1e-10	5.7e-02	4.1e-02	6.6e-02	7.3e-01	2.8e-01
sleep_score_disturbances	3.6e-50	2.2e-02	7.8e-02	1.0e-01	4.8e-01	3.2e-01
sleep_bedtime_start_delta	8.3e-43	9.6e-02	1.1e-02	1.9e-01	4.2e-01	3.0e-01
sleep_temperature_delta	1.5e-54	1.8e-02	1.1e-01	1.5e-01	4.2e-01	3.2e-01

Figure 4.14: The results of COSANOVA for on readiness dataset. The heatmap represents the top 10 features by the variability when calculating VANOVA and HANOVA respectively. The left part is HANOVA, representing the variability between high- and low-level readiness groups. The right part is VANOVA, representing the variability across time windows. The red cells indicate significant variability. A darker red color represents greater significance. (Significance threshold $\sigma < 0.05$)

Case2: Quantifying Rhythm Variability on Real-world Well-being dataset

The dataset utilized smartphone, Fitbit, and daily self-evaluation survey to collect data from 167 participants at an American university over the course of one semester. A commercial device, Fitbit Sense, was used to collect physiological and physical activity data, including steps, calories burned, sleep logs, and heart rates. Smartphone data was collected through the AWARE app. AWARE is an open-source framework [108], including phone calls, screen usage, and location. Table 4.13 gives a full list of all sensor features from Fitbit and AWARE. In addition, and daily survey questions gathered ecological momentary assessment (EMA) information about 10 types of well-

being, including cognitive energy, emotional energy, physical energy, engagement in professional activity, engagement in social activity, engagement in physical activity, mental demand, physical demanding, sleep quality, and rushment. Participants will rate each of these 10 well-being types using numbers from one to five subjectively at the end of every day. The default value of the survey questions is three.

Abbr.	Definition
HR	hear rates
Sleep	sleep levels (e.g., restless, aslepp, awake)
Step	steps
Calorie	burned calories
ACT1	time in the sedentary activity level
ACT2	time in the light actitve activity level
ACT3	time in the moderate activity level
ACT4	time in the very active activity level
CALL1	length of the phone call session
CALL2	phone call types (e.g., incoming, outgoing, missed)
BAT1	phone battery level
BAT2	phone battery voltage
BAT3	phone battery temperature
LOC1	current location's longitude
LOC2	current location's lattitude
LOC3	current location's altitude
LOC4	current location's bearing
LOC5	current location's speed (meters/second)
LOC6	estimated location accuracy
Screen	screen status (e.g., off, on, locked, unlocked)

Table 4.13: Abbreviation and definition of sensor features.

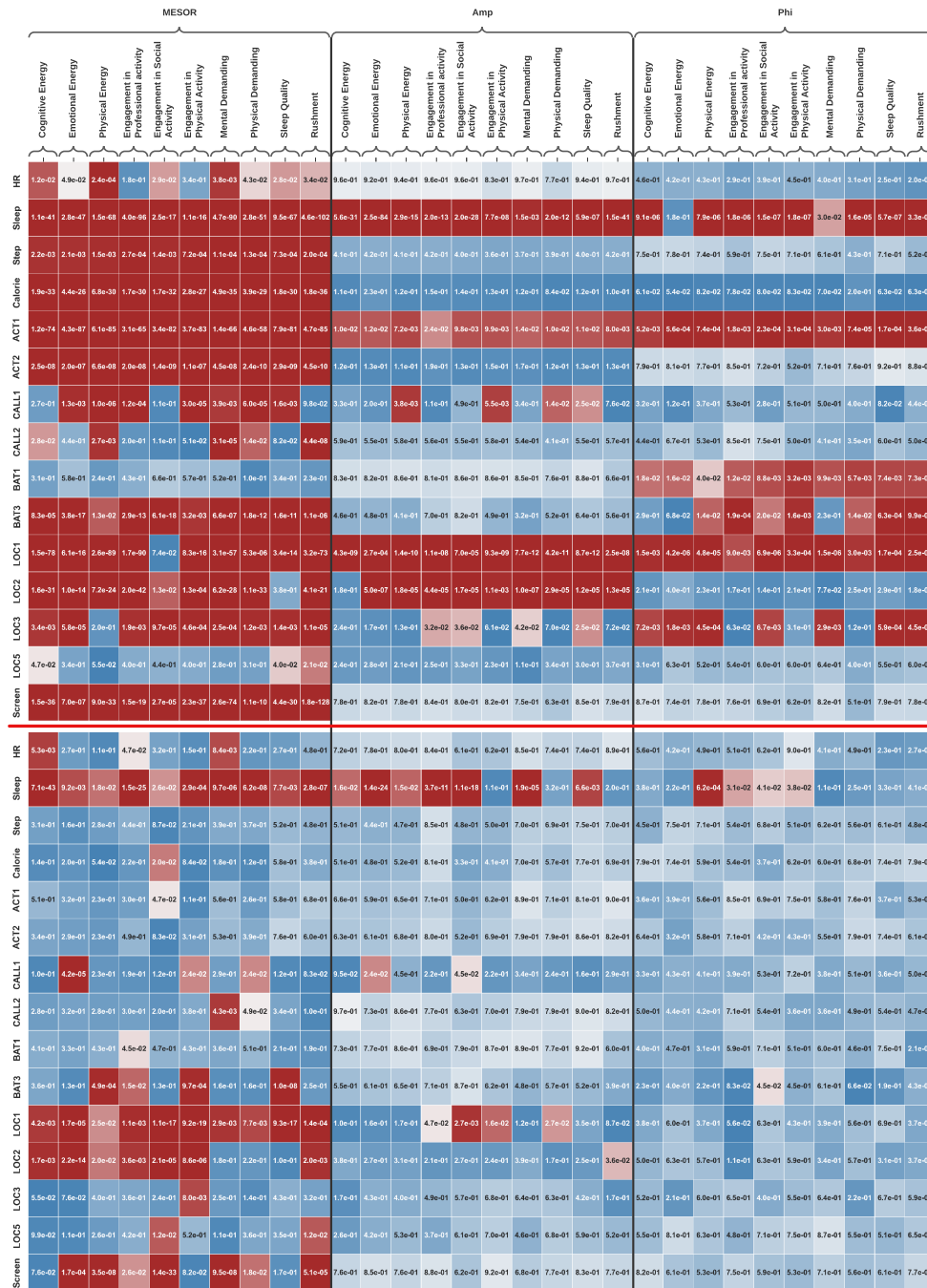


Figure 4.15: The results of COSANOVA for a combination of sensor feature, rhythm parameter and well-being type, and a low value indicates high variability in the rhythm model of a specific sensor feature. The top part is HANOVA, representing the variability between high- and low-level well-being groups. The bottom part is VANOVA, representing the variability across time windows. The red cells indicate significant variability. A darker red color represents greater significance.

To ensure the robustness of the rhythm variability results, I removed participants who had data for less than a complete month. After removal, a total of 120 participants remain. The values of sensing data are assumed to be missing at random (MAR), and I use the simple moving average (SMA) to impute the missing values. SMA replaces the missing values by averaging the non-missing values within a rolling window without weights. I then split the whole semester's data into tumbling time windows of one month.

Figure 4.15 visualizes the results of COSANOVA to show stability of rhythms in each sensor feature between high-level and low-level groups (VANOVA), and across time windows (HANOVA) for each well-being type. A low value shown in the figure indicates high stability in a specific rhythm parameter. The heatmap above the red horizontal line represents the results of VANOVA, while the part below the red line represents the results of HANOVA.

For VANOVA, the rhythm parameter MESOR has smaller values than Amp and Phi, indicating the midline of rhythm models had a more significant difference between high and low-level well-being groups than amplitude and phase. The sensor feature sleep level (Sleep), call duration (CALL1), longitude (LOC1), and latitude (LOC2) of visited location had low values for all three rhythm parameters, and sleep level (Sleep) achieved the generally lowest values of VANOVA in all sensor features. The sensor feature screen had more variability in MESOR but kept relatively stable in Amp and Phi. When specifying well-being type, rhythm models of specific sensor features performed more variability than other features. For example, high- and low-cognitive populations vary greatly in the rhythm models of sleep level (Sleep). This observation agrees that sleep is essential for effective cognitive functioning. High- and low-level engagement in social activity differ significantly in the rhythm models of screen status

(Screen) and phone call duration (CALL1), indicating that students relied on mobile phones for social activities. In addition, the rhythm models of longitude and latitude of the location showed a big difference between high- and low-level engagement in physical activity.

HANOVA had a similar picture of stability scores with VANOVA. MESOR had the most variability across time, and sleep level (Sleep) had the lowest values of HANOVA. However, the phone battery level (BAT1) was most inconsistent in Phi across time, and the movement speed (LOC5) had more fluctuations in Amp across time. The values of HANOVA became smaller than the VANOVA, demonstrating that the rhythm model fluctuates more between different time windows than between different well-being groups. In addition, HANOVA still has some characteristics that are different from VANOVA. For example, the sensor feature related to activity level, such as time spent in the sedentary activity level (ACT1), time spent in the light active activity level (ACT2), and burned calories (Calorie), achieved much smaller values in HANOVA than VANOVA, especially for rhythm parameter. However, Physiological indicator heart rate (HR) had similar HANOVA and VANOVA values, suggesting that HR's rhythm model is relatively stable across populations and time windows. The phone battery level's (BAT1) HANOVA results even rose slightly compared to its VANOVA value.

4.4 Mining Sub-patterns for Human Activity Recognition with WTL

The proposed computational framework integrates the Wavelet Transform Leader (WTL) component (discussed in Section 3.3) to facilitate a comprehensive analysis of human behavior. I assess the performance of the suggested Wavelet Transfer Learning method on real-world sensing data for modeling and interpreting human activities. I evaluate WTL's capacity to identify and infer 1) clusters of similar activity patterns, and 2) various sub-patterns characterized in different regions of the spectrum image. To examine the approach's performance on single- and multi-modal, as well as cyclic and non-cyclic data, I conducted experiments on two human activity datasets: WISDM [83] and PAMAP2 [91]. The WISDM dataset is a single-modal human activity dataset containing accelerometer data from 36 users engaging in six activities: ascending stairs, descending stairs, jogging, sitting, walking, and standing. The raw accelerometer data was collected at a sampling frequency of 20Hz for 3 minutes per activity. The PAMAP2 dataset is a multi-modal human activity dataset that includes accelerometer, gyroscope, heart rate, and skin temperature data collected from nine subjects performing 12 different physical activities. The motion and heart rate sensors have sampling frequencies of 100Hz and 9Hz, respectively. The activities with more regular cycles include walking, running, cycling, Nordic walking, rope jumping, and ascending and descending stairs. The non-cyclic activities consist of lying, sitting, standing, vacuum cleaning, and ironing.

4.4.1 Activity Clustering

Baseline Clustering Models

I compare WTL’s performance to that of six clustering algorithms. The baseline approaches can broadly be divided into two categories **1) time-series clustering algorithms**: K-means [58], Hierarchical Clustering [123], and Self-Organizing Maps [86]), and **2) deep image clustering algorithms**: Deep Embedded Clustering [129], Deep Adaptive Clustering [133], and Semantic Clustering by Adopting Nearest Neighbors [174].

K-means proceeds by alternating between the cluster assignment and the centroid update steps. The shortcoming of k-means is that the value of clusters must be determined beforehand [58]. I use the Elbow method to determine the number of clusters. **Hierarchical Clustering (HC)** algorithm works by grouping time series into a tree of clusters [123]. Hierarchical clustering does not require the number of clusters to be pre-specified.

Self-organizing map (SOM) algorithm is an unsupervised neural network commonly used for high-dimensional data clustering [86]. SOM has a two-dimensional structure of neurons and is trained by an iterative unsupervised self-organizing procedure.

Deep Embedded Clustering (DEC) is a clustering algorithm that combines deep learning and unsupervised clustering techniques [129]. It first pre-trains an autoencoder neural network to reconstruct the input images in a lower-dimensional space. Then it applies the k-means algorithm to the embedded data produced by the encoder. This process is repeated until the autoencoder produces a low-dimensional

representation of the well-suited data for clustering. **Deep Adaptive Clustering (DAC)** is an image clustering algorithm that uses a pairwise binary classification framework. Given two input images, the model outputs whether the inputs belong to the same cluster [133]. **Semantic Clustering by Adopting Nearest Neighbors (SCAN)** is an image clustering algorithm proposed to improve clustering semantics. It includes a two-step procedure that first learns semantic features and then adopts the obtained semantic features as a prior in a deep clustering network [174]). I use the the fused spectrum images as the input to each of these deep image clustering.

To test the performance of the approach on single- and multi-modal as well as cyclic and non-cyclic data, I conducted experiments on two human activity datasets: WISDM [83] and PAMAP2 [91]. The WISDM dataset is a single-modal human activity dataset that contains accelerometer data collected from 36 users performing six activities: ascending stairs, descending stairs, jogging, sitting, walking, and standing. The raw accelerometer data is collected at a sampling frequency of 20Hz for 3 minutes apiece. The PAMAP2 dataset is a multi-modal human activity dataset that includes accelerometer, gyroscope, heart rate, and skin temperature data collected from nine subjects performing 12 different physical activities. The motion and heart rate sensors have sampling frequencies of 100Hz and 9Hz, respectively. The activities with more regular cycles include walking, running, cycling, Nordic walking, rope jumping, and ascending and descending stairs. The non-cyclic activities include lying, sitting, standing, vacuum cleaning, and ironing.

Evaluation Metric

To assess how well the method can cluster human behavior, I divide each participant's spectrum image into M temporal samples. I then cluster each temporal sample using

WTL and six baseline methods. I measure the performance of each approach using the consistency ratio (CR), the ratio of the number of times participants' samples were given the same cluster assignment to the total number of samples in the dataset:

$$CR = \frac{\sum_i^N F(CL_{i1}, \dots, CL_{iM})}{N} \quad (4.1)$$

$$F(CL_{i1}, \dots, CL_{iM}) = \begin{cases} 1 & \text{if } CL_{i1} = \dots = CL_{iM}; \\ 0 & \text{otherwise} \end{cases}$$

, where M is the number of temporal samples obtained from a single spectrum image, CL_{iM} is the cluster label assigned for the M th temporal sample of the i th participant, and N is the total number of participants in the dataset.

To assess each clustering method I use CR (4.1) with $M = 2$. Table 4.14 reports these results. WTL achieves the highest CR across all tested methods. Furthermore, the results demonstrate that the deep image clustering methods (DEC, DAC, SCAN, and WTL) that utilize the Wavelet power spectrum representation outperform the time series clustering algorithms (K-means, HC, and SOM).

Dataset	K-means	HC	SOM	DEC	DAC	SCAN	WTL
WISDM	11.7%	10.4%	13.9%	21.1%	18.2%	21.9%	33.0%
PAMAP2_SC	10.8%	10.6%	5.1%	18.7%	17.2%	16.8%	30.0%
PAMAP2_SNC	7.9%	8.8%	4.2%	14.1%	11.1%	12.1%	21.6%
PAMAP2_MC	21.8%	16.1%	14.0%	31.7%	30.9%	35.7%	45.2%
PAMAP2_MNC	11.3%	12.5%	13.3%	22.7%	19.9%	24.7%	31.9%

Table 4.14: Consistency of clustering across different temporal samples; a high percentage indicates better consistency. The multi-modal human activity dataset PAMAP2 is divided into five parts: single-modal cyclic (SC); SNC single-modal non-cyclic (SNC), multi-modal cyclic (MC), and multi-modal non-cyclic (MNC).

4.4.2 Behavioral Sub-patterns

As previously described, the proposed WTL can be used to inspect behavioral patterns and find sub-patterns. By sub-patterns, I mean individuals act differently even when engaged in the same behavior. Everyone’s activity pattern (e.g., walking style) may be unique and stable over time or change in different conditions and situations. For example, one might walk slowly and steadily for relaxation but fast and inconsistently in a hurry. Both are labeled as walking but may project different patterns that characterize the nature of the walk. The visual inspection of spectrum samples for walking activities in Figure 4.17 shows that individuals can display diverse patterns for the same activity. In Figure 4.16, I show the distribution of the number of pseudo-labels assigned per individual per activity. For example, the figure demonstrates that 12.5% of participants exhibited one ironing sub-pattern, 50% two ironing sub-types, and 37.5% exhibited three ironing sub-types. The figure also shows that most people have multiple behavior patterns for several activities. Specifically, I see most people have at least two behavior patterns for ironing, vacuum cleaning, standing, sitting, Nordic walking, running, and walking. In fact, at least one participant exhibited three distinct behavior patterns for each of these activities. There is only one activity — descending stairs — for which no participant exhibits multiple sub-patterns.

Despite each participant exhibiting somewhere between one and three behavior sub-patterns per activity, as reported in Table 4.15, the patterns are very diverse. For example, I see that 11 examples of ironing were labeled “chain”, 5 were labeled “comic book”, 3 were labeled “beer glass”, 3 were labeled “fountain”, and the 10 remaining examples were assigned distinct labels. This demonstrates that there are many ways in which a human can perform even simple behaviors.

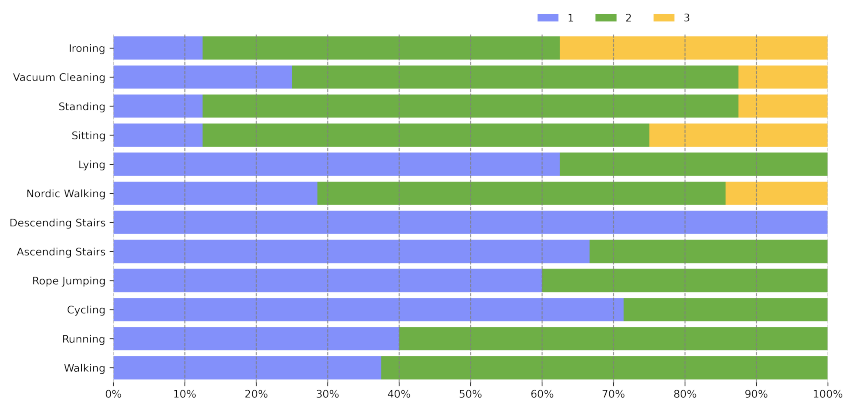


Figure 4.16: The distribution of the number of pseudo-labels assigned per individual per activity. For each activity, the length of the purple bar indicates the proportion of individuals whose spectral images were assigned only one type of pseudo-label. The green represents the proportion of individuals whose spectrum images have two different pseudo-labels, and the yellow represents three different pseudo-labels.

Activity	Pseudo-label Num
Walking	tray: 13, jersey: 5, paintbrush: 4, bucket: 4, binder: 4, others: 3
Running	jersey: 3, comic_book: 2, laptop: 2, others: 4
Cycling	quilt: 7, wall_clock: 5, jersey: 4, chain: 4, others: 3
Rope Jumping	laptop: 2, jersey: 1, barn: 1, bucket: 1, tray: 1, analog_clock: 1
Ascending Stairs	paintbrush: 4, comic_book: 2, bucket: 2, others: 4
Descending Stairs	paintbrush: 3, street_sign: 2, chain: 2, rain_barrel: 2
Nordic Walking	jersey: 11, greenhouse: 4, comic_book: 3, laptop: 3, others: 5
Lying	web_site: 5, mailbox: 5, jersey: 4, laptop: 3, others: 8
Sitting	rain_barrel: 5, jersey: 3, web_site: 3, laptop: 2, others: 8
Standing	padloc: 12, jersey: 7, quill: 4, others: 5
Vacuum Cleaning	fountain: 7, web_site: 6, ashcan: 5, others: 6
Ironing	chain: 11, comic_book: 5, beer_glass: 3, fountain: 3, others: 10

Table 4.15: The table lists the frequency of pseudo-labels in each activity. The most frequent pseudo-labels are ranked first in each activity, and I use "others" to represent those pseudo-labels with low frequency.

4.4.3 Granular Behavior Inspection

As shown in Figure 3.5, the model uses the time window length and frequency band to split a whole Wavelet power spectrum into pieces and model the fine-grained patterns in different regions. In this section, I will demonstrate the existence of distinctive

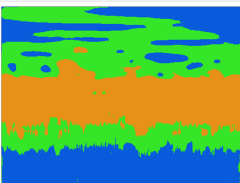
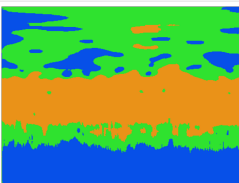
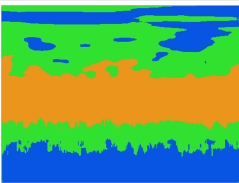
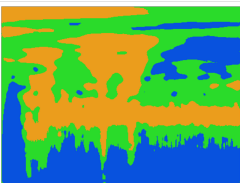
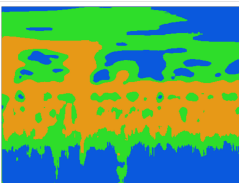

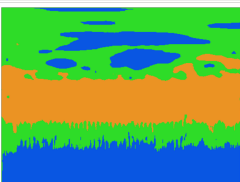
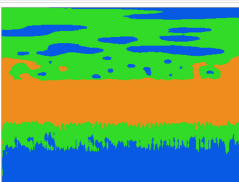
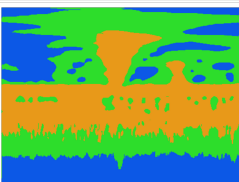


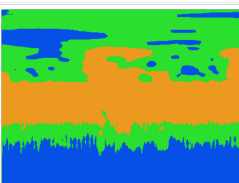
Activity	Participant	Spectrum Image Sample		
Walking	101	tray	tray	tray
				
Walking	104	bucket	bucket	bucket
				
Walking	108	tray	tray	paintbrush
				
Walking	103	jersey	jersey	paintbrush
				

Figure 4.17: The walking spectrum images from the participants: 101, 104, 108, and 103. While the images from participants 101 and 104 were assigned one pseudo-label each, 108 and 103 have two pseudo-labels.

fine-grained patterns in diverse regions of the image, and examine the number of distinct pseudo-labels assigned to the entire spectrum image. Figure 4.18 presents the dominant patterns expressed by pseudo-labels observed in six activities. The fused spectrum images were split into the regions, horizontal and vertical, corresponding to time windows and frequency bands. The idea was to see how consistent the patterns of different areas of the fused spectrum are for each activity. The figure shows that a common pattern exists for the top and bottom regions of the spectrum image for most activities. For example, in 52% of walking samples, the top left and right regions

(frequency domains) and the bottom left and right regions (time domains) of the fused spectrum images have identical pseudo-labels. However, this consistency is observed in less than half of the samples for each activity. This may indicate wide variations in the inferred patterns in each region for different activities caused by differences in activity patterns among people.

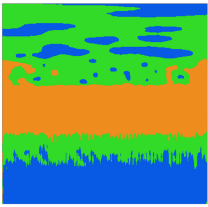

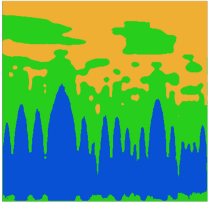
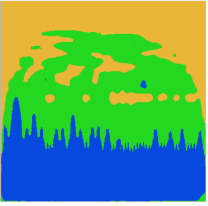
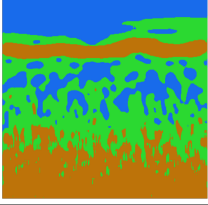
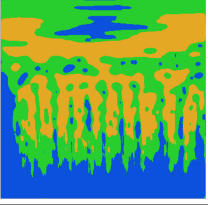
Activity	Processed Spectrum	Perc	Pseudo Labels	Activity	Processed Spectrum	Perc	Pseudo Labels
Walking		52%	"scuba diver", "scuba diver", "envelope", "envelope"	Ironing		24%	"analog clock", "analog clock", "jersey", "jersey"
Sitting		36%	"tray", "balloon", "quill", "quill"	Ascending Stairs		42%	"website", "website", "envelop", "envelop"
Lying		33%	"envelop", "envelop", "mailbox", "mailbox"	Descending Stairs		44%	"tray", "tray", "chain", "chain"

Figure 4.18: Examples of the processed Wavelet power spectrum for six activities and their frequently assigned pseudo-labels. The column of "Perc" presents the percentage of the Wavelet power spectrum obtained from the combination of pseudo-labels shown in the "Pseudo Labels" column. pseudo-labels in each cell are for regions from the top left, top right, bottom left, and bottom right.

The distribution of distinct pseudo-labels assigned per split Wavelet image is presented in Figure 4.19. The figure provides insights into the granular behavior patterns across different regions of the entire spectrum image and shows the number of distinct granular patterns present in one spectrum image. For example, the same pseudo-label was assigned to all the regions of the Wavelet image in 4% of vacuum cleaning examples, two distinct pseudo-labels were assigned to 60% of vacuum clean-

ing examples, three distinct pseudo-labels were assigned to 28% of vacuum cleaning examples, and fthe distinct pseudo-labels were assigned to 8% of vacuum cleaning examples. I can observe that most activities have at least two regions for which the same pseudo-label was assigned. In fact, vacuum cleaning, Nordic walking, and walking are the only activities for which at least one example was assigned the same label for each region. By contrast, all activities had at least one example for which all fthe regions were assigned different labels.

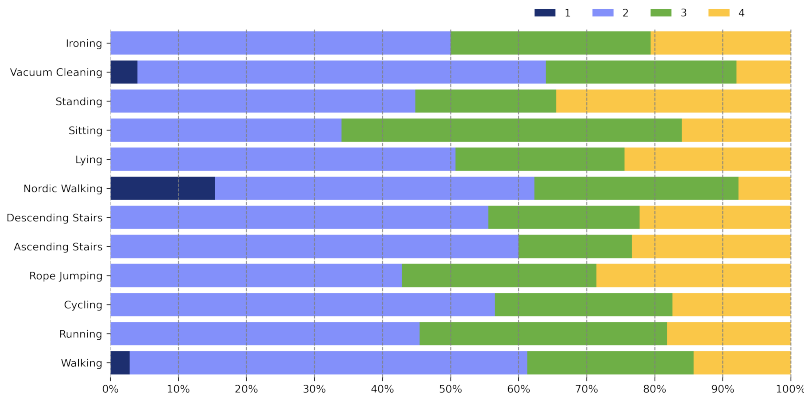


Figure 4.19: The distribution of the distinct pseudo-labels assigned to fthe regions of Wavelet spectrum image. I count the proportion of images given the same pseudo-labels, two different pseudo-labels, three different pseudo-label, and fthe different pseudo-labels, respectively. For most activities, their images have been assigned two different pseudo-labels.

In order to exhibit the proficiency of the proposed computational framework for the evaluation of bibehavioral rhythms derived from multi-modal sensor data, this study introduces fthe distinct applications that leverage constructed rhythm models to investigate human health and well-being outcomes. These applications consist of (1) employing machine learning models for outcome prediction, (2) the discovery of sub-patterns to discern crucial human behavioral patterns, (3) the quantification of rhythm variability, and (4) the examination of relationships between rhythms and

health and well-being outcomes. By utilizing a diverse collection of datasets, the study highlights the framework's flexibility in handling different types of sensor data and measuring various outcomes.

Chapter 5

Discussion

In this thesis, a novel computational framework was developed for modeling cyclic human behaviors using multi-modal sensor data. The primary contribution lies in creating a versatile computational framework capable of analyzing biobehavioral rhythms across a range of sensor data types and time granularities. This framework has potential applications for improving overall well-being.

Initially, I identified key challenges for developing an automated pipeline to process and model biobehavioral time series data from mobile and wearable devices. These challenges include 1) handling and processing vast amounts of multi-modal sensor data, 2) granular examination of signals to extract biobehavioral cycle information, and 3) computational steps for modeling, discovering, and quantifying common patterns.

The implementation involved detecting periods and modeling rhythm shapes by exploring both stationary and non-stationary time series methods. For stationary time series, well-established theories and methods were employed for analyzing periods and rhythm shapes. Three periodicity detection algorithms were used to examine commonality in detecting periods in stationary time series data without ground truth. For non-stationary time series, Wavelet transform (WT) was incorporated to handle non-stationary time series by encoding raw sensor data in an image-based representation, enabling the identification of frequencies and their occurrence times within a

signal.

In modeling rhythmic stability, periodicity detection was first carried out on time series data, followed by an exploration of variations within each period using the changing point detection technique. A novel approach, COSANOVA, was introduced to quantify cyclic variability among different populations and across time windows. This approach demonstrated the ability to infer variability of cyclic time series and determine the contribution of periodic components to the variation. When applied to real-world mobile sensing datasets, it provided insights into the variability of health-related rhythms in various human behavioral and physiological indicators.

To identify relationships between biobehavioral rhythm stability, a correlation-based method was developed to measure multi-dimensional relationships between biobehavioral rhythm stability and productivity across three different weeks. The results showed more rhythm stability in the high-productivity group of students in the sample, despite variations in students' workload in different weeks. Moreover, correlation analysis of rhythms for each sensor feature, along with machine learning models, demonstrated the predictive power of regularity in location-related patterns and sleep.

In predicting health and well-being outcomes, rhythm models were built for each sensor feature, and the generated cyclic parameters were used as inputs for machine learning models to predict outcomes on a per-person and per-time-window basis. Three case studies predicting mental health, readiness score, and productivity were presented to demonstrate the capabilities of the proposed computational framework. The results revealed that the combination of multiple sensor features improved prediction results, and models built with rhythmic features outperformed those built with raw sensor features.

In mining rhythmic patterns, a method combining Wavelet image representation with transfer learning was proposed to enable human interpretation and granular inspection of inferred behaviors. The method was evaluated in terms of inferring activity patterns and labels, providing granular information about behavior patterns, and enabling more granular inspections of behavior patterns that reveal time and frequency-dependent characteristics. Despite the promising performance of the proposed WTL method compared to six baseline clustering algorithms, there are several limitations to consider.

5.1 Limitations

Due to the absence of ground-truth labels for sub-patterns in public datasets, I have proposed the consistency ratio of clustering as an evaluation metric. While it provides a useful measure for comparing clustering algorithms, it does not fully capture the true quality of the clustering results, as it mainly focuses on temporal split consistency rather than the semantic meaning of the sub-patterns or their relevance to the overall physical activity. The consistency ratio is based on the assumption that human behavioral patterns remain stable within each input sample, and therefore, the behavioral pattern of each temporal split should be similar to the others. However, this assumption may not hold true for all scenarios, as human behavior can be dynamic and change within short periods.

The evaluation metric of consistency ratio relies on the temporal splits of input samples, assuming these splits effectively capture the sub-patterns in the data. However, different choices of temporal splits could lead to varying clustering results and impact the consistency ratio, which might affect the overall assessment of the WTL method.

Real-world human activity data often contains noise and artifacts that can significantly impact the quality of extracted sub-patterns. The robustness of the WTL method in dealing with such noisy data and its ability to discern true sub-patterns amidst noise has not been thoroughly examined. The effectiveness of the WTL method has been demonstrated on a specific type of human physical activity data. However, its applicability to other types of activities or different domains has not been established. Additional validation is required to ensure that the WTL method can effectively generalize across a broader range of scenarios.

5.2 Future Work

The proposed WTL approach shown in Figure 3.5 has potential applications in monitoring behavioral health. For instance, it can be employed to identify emotional states or psychological conditions by detecting sub-patterns in human behaviors. One key aspect of recognizing sub-patterns involves detecting minute behavioral changes that may otherwise go unnoticed. This can offer valuable insights into an individual's emotional and psychological health. By identifying unique sub-patterns, mental health professionals can create customized interventions that are specifically designed to address a patient's needs and preferences. Furthermore, determining these sub-patterns allows for more accurate monitoring of a patient's progress, facilitating the evaluation and adjustment of treatment plans as necessary.

To gather data for this study, various aspects of an individual's behavior can be captured using multiple sensors, such as wearable devices that monitor physiological signals, physical activity, and sleep patterns. Additionally, smartphone usage data, including app usage patterns and social interactions, can be analyzed to gain insights

into a person's mental well-being. The WTL method can then be applied to detect sub-patterns in the behavioral data, which may correspond to specific emotional states or psychological conditions. Unsupervised clustering methods can be utilized to group similar patterns. Upon identification of the sub-patterns, mental health experts can analyze their features and evaluate their significance in relation to the individual's mental health. This may include comparing the identified sub-patterns to normative data or existing diagnostic criteria, examining the connections between sub-patterns and self-reported mood evaluations, and analyzing the temporal patterns and consistency of sub-patterns over time. Recognizing behavioral patterns can help mine specific underlying sub-groups and individual patterns within the data, leading to a better understanding of individual differences and promoting the development of personalized interventions or suggestions. As a result, behavioral pattern recognition techniques can adapt to changing environments or evolving behaviors by detecting new or emerging patterns. This flexibility is crucial for applications such as behavioral health monitoring, where a person's behavior may evolve due to interventions, life events, or personal growth.

Chapter 6

Conclusions

In this thesis, the primary contribution is the development of an innovative computational framework to analyze biobehavioral rhythms using diverse sensor data. The framework has the ability to process a wide variety of sensing signals at different time granularities, reveal underlying patterns in human behavior, model the shape and variability of rhythms, and apply these models to predict health outcomes. By providing a practical solution to decipher cyclic human behaviors from sensor data, this framework holds promise for numerous applications aimed at enhancing overall well-being.

One such application is the potential to assist in diagnosing and treating diseases. Medical practitioners can have patients wear sensors, allowing the framework to detect disruptions in biobehavioral rhythms that may be indicative of health issues. By tracking these rhythm changes, doctors can develop a more comprehensive understanding of a patient's condition and create personalized treatment plans to restore rhythmic balance.

Additionally, the framework can help individuals manage their daily schedules more effectively by increasing their awareness of personal biobehavioral rhythms. With this insight, people can make informed decisions about their activities based on their current state. For example, they may allocate time for cognitively demanding tasks when their cognitive rhythm is at its peak and schedule relaxation when their emo-

tional rhythm is at its lowest point. Integrating this framework into existing task management applications, such as Google Calendar, could further support efficient time management and foster a healthy work-life balance.

In conclusion, the computational framework presented in this thesis has considerable potential to improve health outcomes, optimize daily routines, and promote balanced lifestyles. As technology continues to advance, it is expected that applications like this will become increasingly prevalent, enabling individuals to achieve greater productivity and well-being through a deeper understanding of their cyclic human behaviors.

Bibliography

- [1] J Enright. “The search for rhythmicity in biological time-series”. In: *Journal of theoretical biology* 8 (June 1965), pp. 426–68. doi: [10.1016/0022-5193\(65\)90021-4](https://doi.org/10.1016/0022-5193(65)90021-4).
- [2] Franz Halberg et al. “Spectral resolution of low-frequency, small-amplitude rhythms in excreted 17-ketosteroids; probable androgen-induced circaseptan desynchronization”. In: *European Journal of Endocrinology* 50.4_Suppl (1965), S5–S54.
- [3] Franz Halberg, Yung Liang Tong, and EA Johnson. “Circadian system phase—an aspect of temporal morphology; procedures and illustrative examples”. In: *The cellular aspects of biorhythms*. Springer, 1967, pp. 20–48.
- [4] Franz Halberg. “Chronobiology”. In: *Annual review of physiology* 31.1 (1969), pp. 675–726.
- [5] J. Gani and P. Bloomfield. “Fourier Analysis of Time Series: An Introduction”. In: *International Statistical Review / Revue Internationale de Statistique* 46 (Apr. 1978), p. 116. doi: [10.2307/1402516](https://doi.org/10.2307/1402516).
- [6] Genshiro Kitagawa and Hirotugu Akaike. “A procedure for the modeling of non-stationary time series”. In: *Annals of the Institute of Statistical Mathematics* 30 (1978), pp. 351–363.
- [7] Phillip Sokolove and Wayne Bushell. “Sokolove PG, Bushell WN. The chi square periodogram: its utility for analysis of circadian rhythms. *J Theor Biol*

- 72: 131-160". In: *Journal of theoretical biology* 72 (June 1978), pp. 131–60. DOI: [10.1016/0022-5193\(78\)90022-X](https://doi.org/10.1016/0022-5193(78)90022-X).
- [8] John M. Taub. "Behavioral and psychophysiological correlates of irregularity in chronic sleep routines". In: *Biological Psychology* 7.1 (1978), pp. 37–53. ISSN: 0301-0511. DOI: [https://doi.org/10.1016/0301-0511\(78\)90041-8](https://doi.org/10.1016/0301-0511(78)90041-8). URL: <https://www.sciencedirect.com/science/article/pii/0301051178900418>.
- [9] K Abe et al. "Effects of destruction of the suprachiasmatic nuclei on the circadian rhythms in plasma corticosterone, body temperature, feeding and plasma thyrotropin". In: *Neuroendocrinology* 29.2 (1979), pp. 119–131.
- [10] Joseph D Matarazzo. "Computerized clinical psychological test interpretations: Unvalidated plus all mean and no sigma." In: *American Psychologist* 41.1 (1986), p. 14.
- [11] Lawrence Rabiner and B.H. Juang. "An Introduction to Hidden Markov Models". In: *IEEE ASSP Magazine* (Feb. 1986). DOI: [10.1109/MASSP.1986.1165342](https://doi.org/10.1109/MASSP.1986.1165342).
- [12] Lawrence Rabiner and Biinghwang Juang. "An introduction to hidden Markov models". In: *iee assp magazine* 3.1 (1986), pp. 4–16.
- [13] Robert B Cleveland et al. "STL: A seasonal-trend decomposition". In: *J. Off. Stat* 6.1 (1990), pp. 3–73.
- [14] CA Comperatore and GP Krueger. "Circadian rhythm desynchronization, jet lag, shift lag, and coping strategies." In: *Occupational medicine (Philadelphia, Pa.)* 5.2 (1990), pp. 323–341.

- [15] Ronald Aylmer Fisher. “Statistical methods for research workers”. In: *Breakthroughs in statistics*. Springer, 1992, pp. 66–70.
- [16] Roberto Refinetti and Michael Menaker. “The circadian rhythm of body temperature”. In: *Physiology & behavior* 51 (Apr. 1992), pp. 613–37. doi: [10.1016/0031-9384\(92\)90188-8](https://doi.org/10.1016/0031-9384(92)90188-8).
- [17] Pietro Cugini. “Chronobiology: principles and methods”. In: *ANNALI-ISTITUTO SUPERIORE DI SANITA* 29 (1993), pp. 483–483.
- [18] Ping Zhang. “Model selection via multifold cross validation”. In: *The Annals of Statistics* (1993), pp. 299–313.
- [19] Paul M Bentley and JTE McDonnell. “Wavelet transforms: an introduction”. In: *Electronics & communication engineering journal* 6.4 (1994), pp. 175–186.
- [20] S Adamopoulos et al. “Circadian pattern of heart rate variability in chronic heart failure patients Effects of physical training”. In: *European heart journal* 16.10 (1995), pp. 1380–1386.
- [21] Peter Green. “Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination”. In: *Biometrika* 82 (Sept. 1995). doi: [10.1093/biomet/82.4.711](https://doi.org/10.1093/biomet/82.4.711).
- [22] Aaron T Beck et al. “Comparison of Beck Depression Inventories-IA and-II in psychiatric outpatients”. In: *Journal of personality assessment* 67.3 (1996), pp. 588–597.
- [23] Daniel W Russell. “UCLA Loneliness Scale (Version 3): Reliability, validity, and factor structure”. In: *Journal of personality assessment* 66.1 (1996), pp. 20–40.

- [24] Yvonne Chan and Roy P Walmsley. “Learning and understanding the Kruskal-Wallis one-way analysis-of-variance-by-ranks test for differences among three or more independent groups”. In: *Physical therapy* 77.12 (1997), pp. 1755–1761.
- [25] Spyros Makridakis and Michele Hibon. “ARMA models and the Box–Jenkins methodology”. In: *Journal of forecasting* 16.3 (1997), pp. 147–163.
- [26] Rupert G Miller Jr. *Beyond ANOVA: basics of applied statistics*. CRC press, 1997.
- [27] Angela MS Poon et al. “Effect of cage size on ultradian locomotor rhythms of laboratory mice”. In: *Physiology & behavior* 62.6 (1997), pp. 1253–1258.
- [28] David JA Dozois, Keith S Dobson, and Jamie L Ahnberg. “A psychometric evaluation of the Beck Depression Inventory–II.” In: *Psychological assessment* 10.2 (1998), p. 83.
- [29] Alex Pentland and Andrew Liu. “Modeling and prediction of human behavior”. In: *Neural computation* 11.1 (1999), pp. 229–242.
- [30] Elena A Antoniadis et al. “Circadian rhythms, aging and memory”. In: *Behavioural brain research* 111.1-2 (2000), pp. 25–37.
- [31] TG Dietterich. “Ensemble methods in machine learning”. In: Jan. 2000, pp. 1–15. ISBN: 3-540-67704-6.
- [32] Martial M Massin et al. “Circadian rhythm of heart rate and heart rate variability”. In: *Archives of disease in childhood* 83.2 (2000), pp. 179–182.
- [33] Donald B Percival and Andrew T Walden. *Wavelet methods for time series analysis*. Vol. 4. Cambridge university press, 2000.

- [34] M Poyurovsky et al. “Actigraphic monitoring (actigraphy) of circadian locomotor activity in schizophrenic patients with acute neuroleptic-induced akathisia”. In: *European Neuropsychopharmacology* 10.3 (2000), pp. 171–176.
- [35] Leo Breiman. “Random forests”. In: *Machine learning* 45.1 (2001), pp. 5–32.
- [36] Jerome Friedman. “Greedy Function Approximation: A Gradient Boosting Machine”. In: *Annals of Statistics* 29 (Oct. 2001), pp. 1189–1232. doi: [10.2307/2699986](https://doi.org/10.2307/2699986).
- [37] Zoubin Ghahramani. “An introduction to hidden Markov models and Bayesian networks”. In: *International journal of pattern recognition and artificial intelligence* 15.01 (2001), pp. 9–42.
- [38] Gloria Kuhn. “Circadian rhythm, shift work, and emergency medicine”. In: *Annals of emergency medicine* 37.1 (2001), pp. 88–98.
- [39] Rosalind W. Picard, Elias Vyzas, and Jennifer Healey. “Toward machine emotional intelligence: Analysis of affective physiological state”. In: *IEEE transactions on pattern analysis and machine intelligence* 23.10 (2001), pp. 1175–1191.
- [40] Shantha MW Rajaratnam and Josephine Arendt. “Health in a 24-h society”. In: *The Lancet* 358.9286 (2001), pp. 999–1005.
- [41] Yaron Dagan. “Circadian rhythm sleep disorders (CRSD)”. In: *Sleep medicine reviews* 6.1 (2002), pp. 45–55.
- [42] Rob J Hyndman et al. “A state space framework for automatic forecasting using exponential smoothing methods”. In: *International Journal of forecasting* 18.3 (2002), pp. 439–454.

- [43] Kurt Kräuchi. *How is the circadian rhythm of core body temperature regulated?* 2002.
- [44] Scott Menard. *Applied logistic regression analysis*. Vol. 106. Sage, 2002.
- [45] Kenji Yamanishi and Jun-ichi Takeuchi. “A unifying framework for detecting outliers and change points from non-stationary time series data”. In: *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*. 2002, pp. 676–681.
- [46] Marcel Bosc et al. “Automatic change detection in multimodal serial MRI: application to multiple sclerosis lesion evolution”. In: *NeuroImage* 20.2 (2003), pp. 643–656.
- [47] Alain Reinberg and Israel Ashkenazi. “Concepts in human biological rhythms”. In: *Dialogues in clinical neuroscience* 5.4 (2003), p. 327.
- [48] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. “Estimating mutual information”. In: *Phys. Rev. E* 69 (6 June 2004), p. 066138. DOI: [10.1103/PhysRevE.69.066138](https://doi.org/10.1103/PhysRevE.69.066138). URL: <https://link.aps.org/doi/10.1103/PhysRevE.69.066138>.
- [49] Brandon R Peters, Jeff Joireman, and Richard L Ridgway. “Individual differences in the consideration of future consequences scale correlate with sleep habits, sleep quality, and GPA in university students”. In: *Psychological reports* 96.3 (2005), pp. 817–824.
- [50] M Staudacher et al. “A new method for change-point detection developed for on-line analysis of the heart beat variability during sleep”. In: *Physica A: Statistical Mechanics and its Applications* 349.3-4 (2005), pp. 582–596.

- [51] Josephine Arendt. “Melatonin and human rhythms”. In: *Chronobiology international* 23.1-2 (2006), pp. 21–37.
- [52] A Rogier T Donders et al. “A gentle introduction to imputation of missing values”. In: *Journal of clinical epidemiology* 59.10 (2006), pp. 1087–1091.
- [53] Earl F Glynn, Jie Chen, and Arcady R Mushegian. “Detecting periodic patterns in unevenly spaced gene expression time series using Lomb–Scargle periodograms”. In: *Bioinformatics* 22.3 (2006), pp. 310–316.
- [54] Guy P Nason. “Stationary and non-stationary time series”. In: *Statistics in volcanology* 60 (2006).
- [55] Michel F Valstar et al. “Spontaneous vs. posed facial behavior: automatic analysis of brow actions”. In: *Proceedings of the 8th international conference on Multimodal interfaces*. 2006, pp. 162–170.
- [56] Ping Yang, Guy Dumont, and John Mark Ansermino. “Adaptive change detection in heart rate trend monitoring in anesthetized children”. In: *IEEE transactions on biomedical engineering* 53.11 (2006), pp. 2211–2219.
- [57] Timothy I Morgenthaler et al. “Practice parameters for the clinical evaluation and treatment of circadian rhythm sleep disorders”. In: *Sleep* 30.11 (2007), pp. 1445–1459.
- [58] Vit Niennattrakul and Chotirat Ann Ratanamahatana. “On clustering multimedia time series data using k-means and dynamic time warping”. In: *2007 International Conference on Multimedia and Ubiquitous Engineering (MUE’07)*. IEEE. 2007, pp. 733–738.
- [59] Robert L Sack et al. “Circadian rhythm sleep disorders: part I, basic principles, shift work and jet lag disorders”. In: *Sleep* 30.11 (2007), pp. 1460–1483.

- [60] Christina Schmidt et al. “A time to think: circadian rhythms in human cognition”. In: *Cognitive neuropsychology* 24.7 (2007), pp. 755–789.
- [61] Jacopo Aguzzi et al. “A new tracking system for the measurement of diel locomotor rhythms in the Norway lobster, *Nephrops norvegicus* (L.)” In: *Journal of neuroscience methods* 173 (July 2008), pp. 215–24. DOI: [10.1016/j.jneumeth.2008.06.009](https://doi.org/10.1016/j.jneumeth.2008.06.009).
- [62] Pau-Choo Chung and Chin-De Liu. “A daily behavior enabled hidden Markov model for human behavior understanding”. In: *Pattern Recognition* 41.5 (2008), pp. 1572–1580.
- [63] Anne Germain and David Kupfer. “Circadian rhythm disturbances in depression”. In: *Human psychopharmacology* 23 (Oct. 2008), pp. 571–85. DOI: [10.1002/hup.964](https://doi.org/10.1002/hup.964).
- [64] Anne Germain and David J Kupfer. “Circadian rhythm disturbances in depression”. In: *Human Psychopharmacology: Clinical and Experimental* 23.7 (2008), pp. 571–585.
- [65] Rob Hyndman et al. *Forecasting with exponential smoothing: the state space approach*. Springer Science & Business Media, 2008.
- [66] Tom S Price et al. “WAVECLOCK: wavelet analysis of circadian oscillation”. In: *Bioinformatics* 24.23 (2008), pp. 2794–2795.
- [67] Gunilla Brun Sundblad et al. “Self-rated pain and perceived health in relation to stress and physical activity among school-students: A 3-year follow-up”. In: *Pain* 136.3 (2008), pp. 239–249.

- [68] Jia Deng et al. “Imagenet: A large-scale hierarchical image database”. In: *2009 IEEE conference on computer vision and pattern recognition*. Ieee. 2009, pp. 248–255.
- [69] Alex Krizhevsky, Geoffrey Hinton, et al. “Learning multiple layers of features from tiny images”. In: (2009).
- [70] Sinno Jialin Pan and Qiang Yang. “A survey on transfer learning”. In: *IEEE Transactions on knowledge and data engineering* 22.10 (2009), pp. 1345–1359.
- [71] M Sifuzzaman, M Rafiq Islam, and MZ Ali. “Application of wavelet transform and its advantages compared to Fourier transform”. In: (2009).
- [72] Jean-Pierre Etchegaray et al. “Casein kinase 1 delta (CK1 δ) regulates period length of the mouse suprachiasmatic circadian clock in vitro”. In: *PLoS One* 5.4 (2010), e10303.
- [73] Anmol Madan et al. “Social Sensing for Epidemiological Behavior Change”. In: Jan. 2010, pp. 291–300. DOI: [10.1145/1864349.1864394](https://doi.org/10.1145/1864349.1864394).
- [74] Nicolai Meinshausen and Peter Bühlmann. “Stability selection”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 72.4 (2010), pp. 417–473. DOI: [10.1111/j.1467-9868.2010.00740.x](https://doi.org/10.1111/j.1467-9868.2010.00740.x). eprint: <https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9868.2010.00740.x>. URL: <https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2010.00740.x>.
- [75] Lisa Torrey and Jude Shavlik. “Transfer learning”. In: *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*. IGI global, 2010, pp. 242–264.

- [76] Ana Vukolic et al. “Role of mutation of the circadian clock gene *Per2* in cardiovascular circadian rhythms”. In: *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 298.3 (2010), R627–R634.
- [77] Pimwadee Chaovalit et al. “Discrete wavelet transform-based time series analysis and mining”. In: *ACM Computing Surveys (CSUR)* 43.2 (2011), pp. 1–37.
- [78] Alysha M De Livera, Rob J Hyndman, and Ralph D Snyder. “Forecasting time series with complex seasonal patterns using exponential smoothing”. In: *Journal of the American statistical association* 106.496 (2011), pp. 1513–1527.
- [79] Richard Dosselmann and Xue Dong Yang. “A comprehensive assessment of the structural similarity index”. In: *Signal, Image and Video Processing* 5.1 (2011), pp. 81–91.
- [80] John Gale et al. “Disruption of Circadian Rhythms Accelerates Development of Diabetes through Pancreatic Beta-Cell Loss and Dysfunction”. In: *Journal of biological rhythms* 26 (Oct. 2011), pp. 423–33. DOI: [10.1177/0748730411416341](https://doi.org/10.1177/0748730411416341).
- [81] John E Gale et al. “Disruption of circadian rhythms accelerates development of diabetes through pancreatic beta-cell loss and dysfunction”. In: *Journal of biological rhythms* 26.5 (2011), pp. 423–433.
- [82] Tim LM van Kasteren, Gwenn Englebienne, and Ben JA Kröse. “Human activity recognition from wireless sensor network data: Benchmark and software”. In: *Activity recognition in pervasive intelligent environments*. Springer, 2011, pp. 165–186.

- [83] Jennifer R Kwapisz, Gary M Weiss, and Samuel A Moore. “Activity recognition using cell phone accelerometers”. In: *ACM SigKDD Explorations Newsletter* 12.2 (2011), pp. 74–82.
- [84] Javier Ortiz Laguna, Angel García Olaya, and Daniel Borrajo. “A dynamic sliding window approach for activity recognition”. In: *International Conference on User Modeling, Adaptation, and Personalization*. Springer. 2011, pp. 219–230.
- [85] F. Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.
- [86] Peter Sarlin and Tomas Eklund. “Fuzzy clustering of the self-organizing map: some applications on financial time series”. In: *International Workshop on Self-Organizing Maps*. Springer. 2011, pp. 40–50.
- [87] Milad Shokouhi. “Detecting seasonal queries by time-series analysis”. In: *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*. 2011, pp. 1171–1172.
- [88] Ruiping Wang et al. “Maximal linear embedding for dimensionality reduction”. In: *IEEE transactions on pattern analysis and machine intelligence* 33.9 (2011), pp. 1776–1792.
- [89] Junghoon Chae et al. “Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition”. In: *2012 IEEE conference on visual analytics science and technology (VAST)*. IEEE. 2012, pp. 143–152.

- [90] Li Deng. “The mnist database of handwritten digit images for machine learning research [best of the web]”. In: *IEEE signal processing magazine* 29.6 (2012), pp. 141–142.
- [91] Attila Reiss and Didier Stricker. “Introducing a new benchmarked dataset for activity monitoring”. In: *2012 16th international symposium on wearable computers*. IEEE. 2012, pp. 108–109.
- [92] Amy M Bohnert, Julie Wargo Aikins, and Nicole T Arola. “Regrouping: Organized activity involvement and social adjustment across the transition to high school”. In: *New directions for child and adolescent development* 2013.140 (2013), pp. 57–75.
- [93] Lars Buitinck et al. “API design for machine learning software: experiences from the scikit-learn project”. In: *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*. 2013, pp. 108–122.
- [94] Sungjoon Choi, Eunwoo Kim, and Songhwai Oh. “Human behavior prediction for smart homes using deep learning”. In: *2013 IEEE RO-MAN*. IEEE. 2013, pp. 173–179.
- [95] Mingsheng Long et al. “Transfer feature learning with joint distribution adaptation”. In: *Proceedings of the IEEE international conference on computer vision*. 2013, pp. 2200–2207.
- [96] Rakesh Malladi, Giridhar P Kalamangalam, and Behnaam Aazhang. “Online Bayesian change point detection algorithms for segmentation of epileptic activity”. In: *2013 Asilomar Conference on Signals, Systems and Computers*. IEEE. 2013, pp. 1833–1837.

- [97] Minoru Yoshizawa, Wataru Takasaki, and Ren Ohmura. “Parameter exploration for response time reduction in accelerometer-based activity recognition”. In: *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication*. 2013, pp. 653–664.
- [98] Saeed Abdullah et al. “Towards circadian computing: ” early to bed and early to rise” makes some of us unhealthy and sleep deprived”. In: *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing*. 2014, pp. 673–684.
- [99] Germaine Cornelissen. “Cosinor-based rhythmometry”. In: *Theoretical Biology and Medical Modelling* 11.1 (2014), pp. 1–24.
- [100] Yipeng Ding and Jingtian Tang. “Micro-Doppler trajectory estimation of pedestrians using a continuous-wave radar”. In: *IEEE Transactions on Geoscience and Remote Sensing* 52.9 (2014), pp. 5807–5819.
- [101] Jun-Ki Min et al. “Toss’n’turn: smartphone as sleep and sleep quality detector”. In: *Proceedings of the SIGCHI conference on human factors in computing systems*. 2014, pp. 477–486.
- [102] M Murugappan, Subbulakshmi Murugappan, Celestin Gerard, et al. “Wireless EEG signals based neuromarketing system using Fast Fourier Transform (FFT)”. In: *2014 IEEE 10th international colloquium on signal processing and its applications*. IEEE. 2014, pp. 25–30.
- [103] Karen Simonyan and Andrew Zisserman. “Very deep convolutional networks for large-scale image recognition”. In: *arXiv preprint arXiv:1409.1556* (2014).

- [104] Rui Wang et al. “StudentLife: Assessing mental health, academic performance and behavioral trends of college students using smartphones”. In: Sept. 2014. doi: [10.1145/2632048.2632054](https://doi.org/10.1145/2632048.2632054).
- [105] Tomasz Zielinski et al. “Strengths and limitations of period estimation methods for circadian data”. In: *PloS one* 9.5 (2014), e96462.
- [106] Frank Bellivier et al. “Sleep-and circadian rhythm-associated pathways as therapeutic targets in bipolar disorder”. In: *Expert opinion on therapeutic targets* 19.6 (2015), pp. 747–763.
- [107] Denzil Ferreira, Vassilis Kostakos, and Anind Dey. “AWARE: Mobile Context Instrumentation Framework”. In: *Frontiers in ICT* 2 (May 2015). doi: [10.3389/fict.2015.00006](https://doi.org/10.3389/fict.2015.00006).
- [108] Denzil Ferreira, Vassilis Kostakos, and Anind K Dey. “AWARE: mobile context instrumentation framework”. In: *Frontiers in ICT* 2 (2015), p. 6.
- [109] Andrew T Jebb et al. “Time series analysis for psychological research: examining and forecasting change”. In: *Frontiers in psychology* 6 (2015), p. 727.
- [110] Elizabeth L Murnane et al. “Social (media) jet lag: How usage of social technology can modulate and reflect circadian rhythms”. In: *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 2015, pp. 843–854.
- [111] Serene S Paul et al. “Validity of the Fitbit activity tracker for measuring steps in community-dwelling older adults”. In: *BMJ open sport & exercise medicine* 1.1 (2015), e000013.

- [112] Sohrab Saeb et al. “Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study”. In: *Journal of Medical Internet Research* 17 (July 2015). doi: [10.2196/jmir.4273](https://doi.org/10.2196/jmir.4273).
- [113] Zhiguang Wang and Tim Oates. “Encoding time series as images for visual inspection and classification using tiled convolutional neural networks”. In: *Workshops at the twenty-ninth AAAI conference on artificial intelligence*. 2015.
- [114] Zhiguang Wang and Tim Oates. “Imaging time-series to improve classification and imputation”. In: *Twenty-Fourth International Joint Conference on Artificial Intelligence*. 2015.
- [115] Zhiguang Wang and Tim Oates. “Spatially encoding temporal correlations to classify temporal data using convolutional neural networks”. In: *arXiv preprint arXiv:1509.07481* (2015).
- [116] Saeed Abdullah et al. “Cognitive rhythms: unobtrusive and continuous sensing of alertness using a mobile phone”. In: *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 2016, pp. 178–189.
- [117] Forest Agostinelli et al. “What time is it? Deep learning approaches for circadian rhythms”. In: *Bioinformatics* 32.12 (2016), pp. i8–i17.
- [118] Mustafa Akpınar and Nejat Yumusak. “Year ahead demand forecast of city natural gas using seasonal time series methods”. In: *Energies* 9.9 (2016), p. 727.
- [119] Mohammad Abu Alsheikh et al. “Deep activity recognition models with tri-axial accelerometers”. In: *Workshops at the Thirtieth AAAI Conference on Artificial Intelligence*. 2016.

- [120] Giannina J Bellone et al. “Comparative analysis of actigraphy performance in healthy young subjects”. In: *Sleep Science* 9.4 (2016), pp. 272–279.
- [121] LAURIE FRICK. *Heart Rate and Skin Temperature Datasets*. <http://https://data.world/laurie/skin-temperature/>. Oct. 2016.
- [122] Kaiming He et al. “Deep residual learning for image recognition”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.
- [123] Maciej Łuczak. “Hierarchical clustering of time series data with parametric derivative dynamic time warping”. In: *Expert Systems with Applications* 62 (2016), pp. 116–130.
- [124] Janna Mantua, Nickolas Gravel, and Rebecca Spencer. “Reliability of sleep measures from four personal health monitoring devices compared to research-based actigraphy and polysomnography”. In: *Sensors* 16.5 (2016), p. 646.
- [125] Elizabeth L Murnane et al. “Mobile manifestations of alertness: Connecting biological rhythms with patterns of smartphone app use”. In: *Proceedings of the 18th international conference on human-computer interaction with mobile devices and services*. 2016, pp. 465–477.
- [126] Christoph Saner et al. “Circadian and ultradian cardiovascular rhythmicity in obese children”. In: *European journal of pediatrics* 175.8 (2016), pp. 1031–1038.
- [127] Christian Szegedy et al. “Rethinking the inception architecture for computer vision”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 2818–2826.
- [128] Karl Weiss, Taghi M Khoshgoftaar, and DingDing Wang. “A survey of transfer learning”. In: *Journal of Big data* 3.1 (2016), pp. 1–40.

- [129] Junyuan Xie, Ross Girshick, and Ali Farhadi. “Unsupervised deep embedding for clustering analysis”. In: *International conference on machine learning*. PMLR. 2016, pp. 478–487.
- [130] Gregory William Yeutter. *Determination of Circadian Rhythms in Consumer-Grade Actigraphy Devices*. Drexel University, 2016.
- [131] Saeed Abdullah et al. “Circadian computing: sensing, modeling, and maintaining biological rhythms”. In: *Mobile health*. Springer, 2017, pp. 35–58.
- [132] Samaneh Aminikhangahi and Diane J Cook. “A survey of methods for time series change point detection”. In: *Knowledge and information systems* 51.2 (2017), pp. 339–367.
- [133] Jianlong Chang et al. “Deep adaptive image clustering”. In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 5879–5887.
- [134] John Cristian Borges Gamboa. “Deep learning for time-series analysis”. In: *arXiv preprint arXiv:1701.01887* (2017).
- [135] DG Gubin et al. “Activity, sleep and ambient light have a different impact on circadian blood pressure, heart rate and body temperature rhythms”. In: *Chronobiology international* 34.5 (2017), pp. 632–649.
- [136] Johnni Hansen. “Night shift work and risk of breast cancer”. In: *Current environmental health reports* 4.3 (2017), pp. 325–339.
- [137] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. “Imagenet classification with deep convolutional neural networks”. In: *Communications of the ACM* 60.6 (2017), pp. 84–90.

- [138] Hyun-Ah Lee et al. “Comparison of wearable activity tracker with actigraphy for sleep evaluation and circadian rest-activity rhythm measurement in healthy young adults”. In: *Psychiatry investigation* 14.2 (2017), p. 179.
- [139] Steffen Moritz and Thomas Bartz-Beielstein. “imputeTS: time series missing value imputation in R.” In: *R J.* 9.1 (2017), p. 207.
- [140] Haya Alaskar. “Deep learning-based model architecture for time-frequency images analysis”. In: *Int. J. Adv. Comput. Sci. Appl* 9.12 (2018), pp. 1–9.
- [141] Ali Ben Abbes et al. “Comparative study of three satellite image time-series decomposition methods for vegetation change detection”. In: *European Journal of Remote Sensing* 51.1 (2018), pp. 607–615.
- [142] Afsaneh Doryab et al. “Extraction of Behavioral Features from Smartphone and Wearable Data”. In: (Dec. 2018).
- [143] Afsaneh Doryab et al. “Extraction of behavioral features from smartphone and wearable data”. In: *arXiv preprint arXiv:1812.10394* (2018).
- [144] Nima Hatami, Yann Gavet, and Johan Debayle. “Classification of time-series images using deep convolutional neural networks”. In: *Tenth international conference on machine vision (ICMV 2017)*. Vol. 10696. International Society for Optics and Photonics. 2018, 106960Y.
- [145] Mary T Imboden et al. “Comparison of four Fitbit and Jawbone activity monitors with a research-grade ActiGraph accelerometer for estimating physical activity and energy expenditure”. In: *British Journal of Sports Medicine* 52.13 (2018), pp. 844–850.

- [146] Brandon R Peters. *Why Does My Heart Rate Spike When I'm Asleep?* 2018. URL: <https://www.beddrsleap.com/posts/why-does-my-heart-rate-spike-when-im-asleep> (visited on 10/10/2020).
- [147] Emma Pierson, Tim Althoff, and Jure Leskovec. “Modeling Individual Cyclic Variation in Human Behavior”. In: *Proceedings of the 2018 World Wide Web Conference*. WWW '18. Lyon, France: International World Wide Web Conferences Steering Committee, 2018, pp. 107–116. ISBN: 9781450356398. DOI: [10.1145/3178876.3186052](https://doi.org/10.1145/3178876.3186052). URL: <https://doi.org/10.1145/3178876.3186052>.
- [148] Emma Pierson, Tim Althoff, and Jure Leskovec. “Modeling individual cyclic variation in human behavior”. In: *Proceedings of the 2018 World Wide Web Conference*. 2018, pp. 107–116.
- [149] Sharif Atique et al. “Forecasting of total daily solar energy generation using ARIMA: A case study”. In: *2019 IEEE 9th annual computing and communication workshop and conference (CCWC)*. IEEE. 2019, pp. 0114–0119.
- [150] Afsaneh Doryab et al. “Identifying behavioral phenotypes of loneliness and social isolation with passive sensing: statistical analysis, data mining and machine learning of smartphone and fitbit data”. In: *JMIR mHealth and uHealth* 7.7 (2019), e13209.
- [151] Afsaneh Doryab et al. “Modeling Biobehavioral Rhythms with Passive Sensing in the Wild: A Case Study to Predict Readmission Risk after Pancreatic Surgery”. In: *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3.1 (Mar. 2019). DOI: [10.1145/3314395](https://doi.org/10.1145/3314395). URL: <https://doi.org/10.1145/3314395>.

- [152] Afsaneh Doryab et al. “Modeling Biobehavioral Rhythms with Passive Sensing in the Wild: A Case Study to Predict Readmission Risk after Pancreatic Surgery”. In: *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3.1 (Mar. 2019). DOI: [10.1145/3314395](https://doi.org/10.1145/3314395). URL: <https://doi.org/10.1145/3314395>.
- [153] Afsaneh Doryab et al. “Modeling biobehavioral rhythms with passive sensing in the wild: a case study to predict readmission risk after pancreatic surgery”. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3.1 (2019), pp. 1–21.
- [154] Quentin Geissmann et al. “Rethomics: An R framework to analyse high-throughput behavioural data”. In: *PloS one* 14.1 (2019).
- [155] Beniamino Hadj-Amar et al. “Bayesian Model Search for Nonstationary Periodic Time Series”. In: *Journal of the American Statistical Association* (May 2019), pp. 1–36. DOI: [10.1080/01621459.2019.1623043](https://doi.org/10.1080/01621459.2019.1623043).
- [156] Hassan Ismail Fawaz et al. “Deep learning for time series classification: a review”. In: *Data mining and knowledge discovery* 33.4 (2019), pp. 917–963.
- [157] Fayez Lahoud and Sabine Süssstrunk. “Zero-learning fast medical image fusion”. In: *2019 22th International Conference on Information Fusion (FUSION)*. IEEE, 2019, pp. 1–8.
- [158] Aileen Nielsen. *Practical time series analysis: Prediction with statistics and machine learning*. O’Reilly Media, 2019.
- [159] Manel Rhif et al. “Wavelet transform application for/in non-stationary time-series analysis: a review”. In: *Applied Sciences* 9.7 (2019), p. 1345.

- [160] Mark O Riedl. “Human-centered artificial intelligence and machine learning”. In: *Human Behavior and Emerging Technologies* 1.1 (2019), pp. 33–36.
- [161] Mingxing Tan and Quoc Le. “Efficientnet: Rethinking model scaling for convolutional neural networks”. In: *International conference on machine learning*. PMLR. 2019, pp. 6105–6114.
- [162] Amin Ullah et al. “Action recognition using optimized deep autoencoder and CNN for surveillance data streams of non-stationary environments”. In: *Future Generation Computer Systems* 96 (2019), pp. 386–397.
- [163] Chao-Lung Yang et al. “Multivariate time series data transformation for convolutional neural network”. In: *2019 IEEE/SICE International Symposium on System Integration (SII)*. IEEE. 2019, pp. 188–192.
- [164] Ozal Yildirim et al. “Automated detection of diabetic subject using pre-trained 2D-CNN models with frequency spectrum images extracted from heart rate signals”. In: *Computers in biology and medicine* 113 (2019), p. 103387.
- [165] Yash Chandak et al. “Optimizing for the future in non-stationary mdps”. In: *International Conference on Machine Learning*. PMLR. 2020, pp. 1414–1425.
- [166] Adam L Halberstadt. “Automated detection of the head-twitch response using wavelet scalograms and a deep convolutional neural network”. In: *Scientific reports* 10.1 (2020), pp. 1–12.
- [167] Alexander Hoelzemann and Kristof Van Laerhoven. “Digging deeper: towards a better understanding of transfer learning for human activity recognition”. In: *Proceedings of the 2020 International Symposium on Wearable Computers*. 2020, pp. 50–54.

- [168] Dana Hughes et al. “Inferring non-stationary human preferences for human-agent teams”. In: *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE. 2020, pp. 1178–1185.
- [169] Wei-Chao Lin and Chih-Fong Tsai. “Missing value imputation: a review and analysis of the literature (2006–2017)”. In: *Artificial Intelligence Review* 53 (2020), pp. 1487–1509.
- [170] Hamid Mansoor et al. “ARGUS: Interactive Visual Analytics Framework for the Discovery of Disruptions in Bio-Behavioral Rhythms.” In: *EuroVis (Short Papers)*. 2020, pp. 25–29.
- [171] Zhen Qin et al. “Imaging and fusing time series for wearable sensor-based human activity recognition”. In: *Information Fusion* 53 (2020), pp. 80–87.
- [172] Antônio H Ribeiro et al. “Automatic diagnosis of the 12-lead ECG using a deep neural network”. In: *Nature communications* 11.1 (2020), pp. 1–9.
- [173] Wei Sun et al. “Vibrosense: Recognizing home activities by deep learning subtle vibrations on an interior surface of a house from a single point using laser doppler vibrometry”. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4.3 (2020), pp. 1–28.
- [174] Wouter Van Gansbeke et al. “Scan: Learning to classify images without labels”. In: *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part X*. Springer. 2020, pp. 268–285.
- [175] Kieran Woodward, Eiman Kanjo, and Athanasios Tsanas. “Combining Deep Transfer Learning with Signal-image Encoding for Multi-Modal Mental Well-being Classification”. In: *arXiv preprint arXiv:2012.03711* (2020).

- [176] Fuzhen Zhuang et al. “A comprehensive survey on transfer learning”. In: *Proceedings of the IEEE* 109.1 (2020), pp. 43–76.
- [177] Sungtae An et al. “AdaptNet: Human Activity Recognition via Bilateral Domain Adaptation Using Semi-Supervised Deep Translation Networks”. In: *IEEE Sensors Journal* 21.18 (2021), pp. 20398–20411.
- [178] Vivaswath S Ayyar and Siddharth Sukumaran. “Circadian rhythms: Influence on physiology, pharmacology, and therapeutic interventions”. In: *Journal of Pharmacokinetics and Pharmacodynamics* 48 (2021), pp. 321–338.
- [179] Michael Borenstein et al. *Introduction to meta-analysis*. John Wiley & Sons, 2021.
- [180] Wei Fang, Yupeng Chen, and Qiongying Xue. “Survey on research of RNN-based spatio-temporal sequence prediction algorithms”. In: *Journal on Big Data* 3.3 (2021), p. 97.
- [181] Da Un Jeong and Ki Moo Lim. “Combined deep CNN–LSTM network-based multitasking learning architecture for noninvasive continuous blood pressure estimation using difference in ECG-PPG features”. In: *Scientific Reports* 11.1 (2021), p. 13539.
- [182] Bryan Lim and Stefan Zohren. “Time-series forecasting with deep learning: a survey”. In: *Philosophical Transactions of the Royal Society A* 379.2194 (2021), p. 20200209.
- [183] Fateme Nikseresht et al. “Detection of Racial Bias from Physiological Responses”. In: *Advances in Usability, User Experience, Wearable and Assistive Technology: Proceedings of the AHFE 2021 Virtual Conferences on Usability and User Experience, Human Factors and Wearable Technologies, Human*

Factors in Virtual Environments and Game Design, and Human Factors and Assistive Technology, July 25-29, 2021, USA. Springer. 2021, pp. 59–66.

- [184] Jonathon Pye et al. “Irregular sleep-wake patterns in older adults with current or remitted depression”. In: *Journal of Affective Disorders* 281 (2021), pp. 431–437.
- [185] Yuan Tian, Ke Zhou, and Dan Pelleg. “What and how long: Prediction of mobile app engagement”. In: *ACM Transactions on Information Systems (TOIS)* 40.1 (2021), pp. 1–38.
- [186] José F Torres et al. “Deep learning for time series forecasting: a survey”. In: *Big Data* 9.1 (2021), pp. 3–21.
- [187] Congyu Wu et al. “Multi-modal data collection for measuring health, behavior, and living environment of large-scale participant cohorts”. In: *GigaScience* 10.6 (2021), giab044.
- [188] Runze Yan and Afsaneh Doryab. “Towards a Computational Framework for Automated Discovery and Modeling of Biological Rhythms from Wearable Data Streams”. In: *Proceedings of SAI Intelligent Systems Conference*. Springer. 2021, pp. 643–661.
- [189] Mingyi Cai, Runze Yan, and Afsaneh Doryab. “Daily Trajectory Prediction Using Temporal Frequent Pattern Tree”. In: *Proceedings of Sixth International Congress on Information and Communication Technology*. Springer. 2022, pp. 333–343.
- [190] Alejandro Sanchez Guinea, Mehran Sarabchian, and Max Mühlhäuser. “Improving Wearable-Based Activity Recognition Using Image Representations”. In: *Sensors* 22.5 (2022), p. 1840.

- [191] Nur Muhammad Shafiullah et al. “Behavior transformers: Cloning k modes with one stone”. In: *Advances in neural information processing systems* 35 (2022), pp. 22955–22968.
- [192] Hao Sun, Yen-Wei Chen, and Lanfen Lin. “TensorFormer: A Tensor-based Multimodal Transformer for Multimodal Sentiment Analysis and Depression Detection”. In: *IEEE Transactions on Affective Computing* (2022).
- [193] Qingsong Wen et al. “Transformers in time series: A survey”. In: *arXiv preprint arXiv:2202.07125* (2022).
- [194] Runze Yan et al. “A Computational Framework for Modeling Biobehavioral Rhythms from Mobile and Wearable Data Streams”. In: *ACM Transactions on Intelligent Systems and Technology (TIST)* 13.3 (2022), pp. 1–27.
- [195] Runze Yan et al. “Exploratory machine learning modeling of adaptive and maladaptive personality traits from passively sensed behavior”. In: *Future Generation Computer Systems* 132 (2022), pp. 266–281.
- [196] *Empatica*. URL: <https://www.empatica.com>.

Appendices

Appendix A

Evaluation of the State-of-the-Art Periodicity Detection Methods

The performance of three fundamentally distinct periodicity detection methods (presented in Section 3.2.1), namely Fast Fourier Transform (FFT), Chi-squared based periodogram, and Cyclic Hidden Markov Models (CyHMMs), will be evaluated. The first dataset is artificially created with known periods to evaluate the effectiveness of the three methods in detecting periods in the data. Subsequently, the methods are examined on a real-world dataset that includes heart rate and temperature data spanning 70 days, as well as a dataset derived from the E4 wearable device. The last dataset comprises 16 days of fine-grained heart rate (HR), heart rate variability (BVP), skin temperature, and galvanic skin response (EDA) data.

A.1 Datasets

I then use three different datasets to evaluate the effectiveness of the above methods. I process the three datasets in the following way:

- Dataset 1 - A synthetic dataset with known periods of 24, 36, and 48 hours (Figure A.1). The sequences are comprised of different sinusoidal signals with

a predefined frequency. I also add 3dB white noise to the dataset to simulate real conditions.

- Dataset 2 - A real-world open-source dataset [121] containing 70 consecutive days of heart rate and skin temperature collected in the one-minute interval as visualized in Figure A.2. The values assumed to be missing at random (MAR) account for 7.63% of the whole dataset. I use the simple moving average (SMA) to impute the missing values [139]. SMA replaces the missing values by averaging the non-missing values within a rolling window without weights.
- Dataset 3 - A real-world dataset collected from Empatica E4 wristband, a medical-grade physiological monitoring device [196], for over two weeks. The E4 device monitors the blood volume pulse (BVP), the electrodermal activity (EDA), the heart rate, and skin temperature in real-time with a sampling rate of 64Hz, 4Hz, 1Hz, and 4Hz respectively. I apply the same imputation method for processing data as in the second dataset.

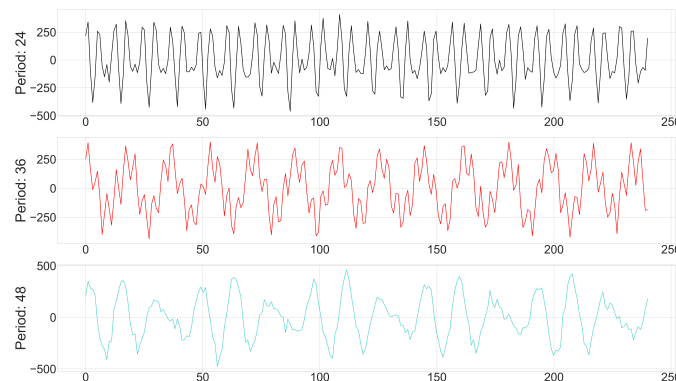


Figure A.1: Visual inspection of the synthetic dataset

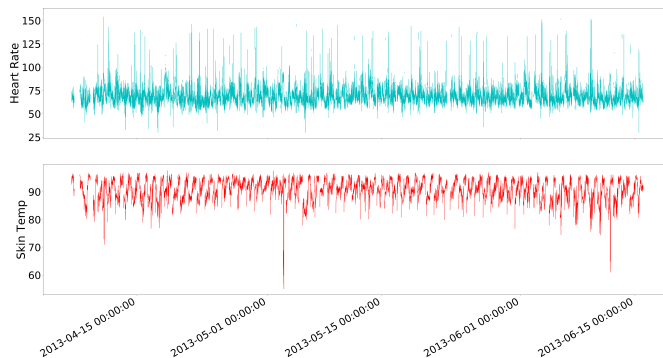


Figure A.2: Visual inspection of the open-source dataset with heart rate and skin temperature data over 70 days

A.2 Periodicity Detection

As shown in Table A.1, all three methods can detect the periods of the synthetic dataset accurately with an error range between 0.09 and 0.36 hours. Among the three methods, the average error of FFT is the smallest, but the difference between FFT and the other two methods is small. These results verify the reliability of the three periodicity detection algorithms.

Synthetic Period	Algorithm	Estimated	Error
24	FFT	23.91	0.09
	Chi	23.89	0.11
	CyHMMs	23.64	0.36
36	FFT	35.72	0.28
	Chi	35.81	0.19
	CyHMMs	35.69	0.31
48	FFT	47.84	0.16
	Chi	47.79	0.21
	CyHMMs	47.87	0.13

Table A.1: The performance of three period estimation algorithms on a synthetic dataset. The results from the synthetic dataset verify the reliability of the three periodicity detection algorithms.

For the second dataset with a 70-day heart rate and temperature, the periodograms outputted by FFT and Chi-square are shown in Figures A.3 and A.4. In the Fourier

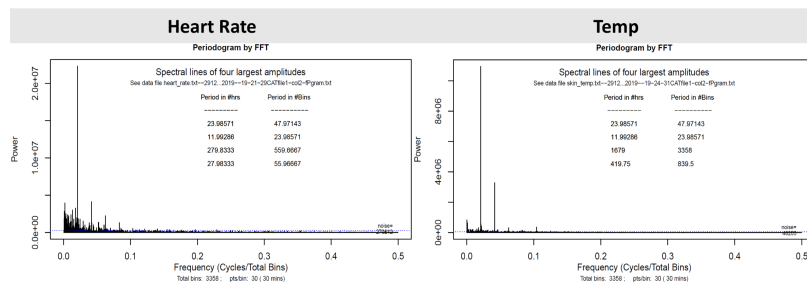


Figure A.3: Fourier periodogram derived from heart rate, and skin temperature for the second dataset. The dashed line indicates the 0.05 level of significance for the periodogram. Dominant frequency correspond to 24 and 12 h for both heart rate and skin temperature.

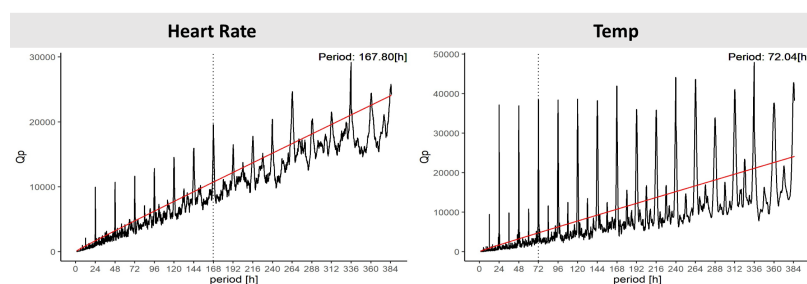


Figure A.4: Chi-square periodograms derived from heart rate, and skin temperature for the second dataset. Red straight lines indicate the significance level of $p = 0.01$. Both heart rate and skin temperature in Chi-square periodogram exhibits a 24 h oscillation.

periodogram, the period according to the dominant frequencies are around 24 and 12 hours. As for the Chi-square periodogram, the most significant periods for heart rate and skin temperature are 167.80 hours and 72.04 h, respectively. However, 168 h and 72 h are both multiples of 12 and 24 h, and the Chi-square periodogram also shows significant oscillations with a period of 24 hours. CyHMMs only infer a period of almost 24 hours for both time series (23.94 h and 24.01 h). Therefore, I can confirm that both heart rate and skin temperature have 24 h rhythms, which is consistent with the circadian cycle. Table A.2 summarizes the detected results by the three methods.

Physiological Data	Algorithm	Detected Periods
Heart Rate	FFT	23.98, 11.99, 279.83, 27.98
	Chi	167.80, 24.00, 12.00
	CyHMMs	23.94
Skin Temperature	FFT	23.98, 11.99, 1679, 419.75
	Chi	72.04, 24.00, 12.00
	CyHMMs	24.01

Table A.2: Periods of heart rate and skin temperature detected by FFT, Chi-square periodogram, and CyHMMs. Heart rate and skin temperature have obvious cycle characteristics of 24 h and 12 h.

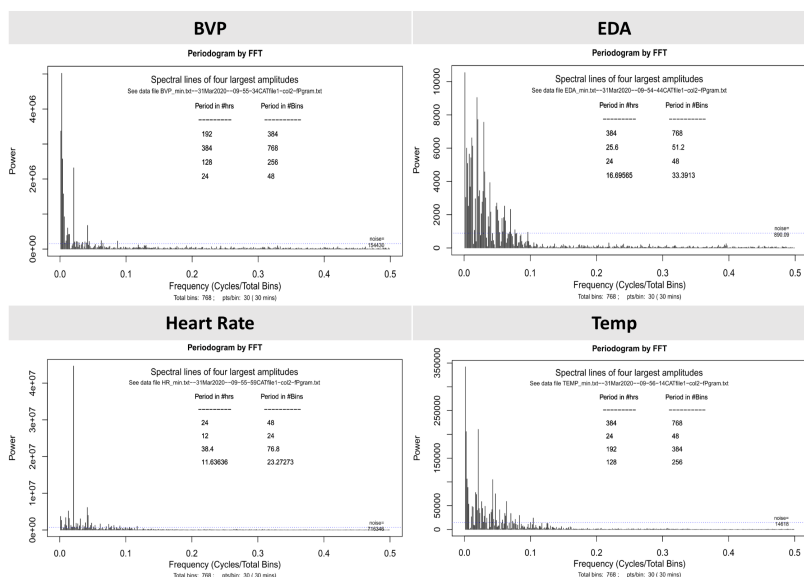


Figure A.5: Fourier periodogram derived from Empatica E4 wristband dataset. The dashed line indicates the 0.05 level of significance for the periodogram. Dominant frequency in BVP, heart rate and skin temperature corresponds to 24 h, or the integral multiple of 24 h. Period corresponding to main frequency in EDA is 25.6 h instead of 24 h. (384 is the total length of the dataset)

Compared to dataset 2, the E4 dataset contains new types of data, namely BVP and EDA, and the data is more fine-grained and clean. In terms of heart rate and skin temperature, FFT and Chi-square periodogram can display a period of 24 h, but the detected period is less significant than dataset 2, which may be a result of shorter time series (about 2 weeks only) and fine-grained high-frequency sampling rates that

result in more fluctuations in the data.

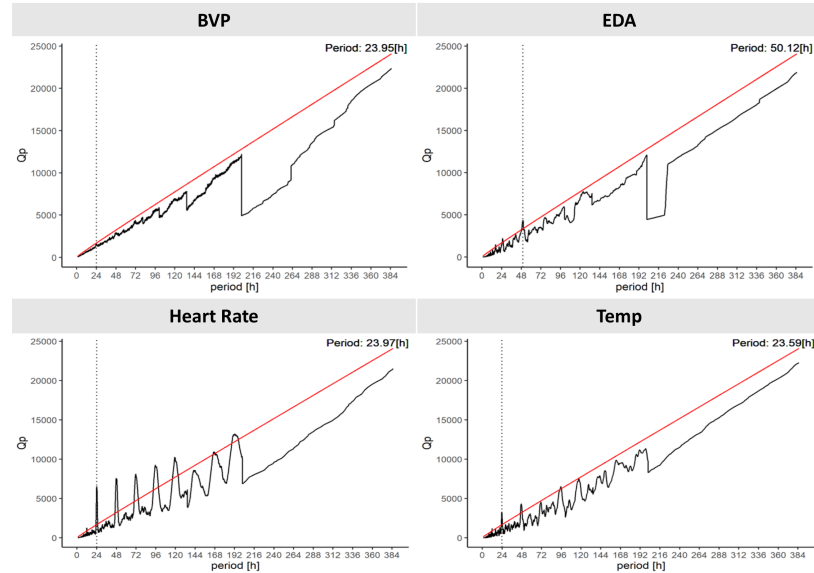


Figure A.6: Chi-square periodograms derived from Empatica E4 wristband dataset. Red straight lines indicate the significance level of $p = 0.01$. Detected period in BVP is 23.95 h. Detected period in EDA is 50.12, and there is a peak in the periodogram at around 25h. For heart rate and skin temperature, the periodograms display a 24 h oscillation.

In contrast to heart rate and skin temperature, the BVP and EDA show less significant periodic patterns. In Fig A.5, BVP has a 24 h period, and EDA is detected to have a 25.6 h period more significant than the 24 h. In Fig A.6, the Chi-square periodogram recognizes that the period of BVP is 23.95 h, which is close to the result derived from the FFT. The rhythm period of EDA is 50.12 h, but in the Chi-square periodogram of EDA, there also exists a peak Q_P value above the significant level at around 25 hours, and this value corresponds to the 25.6 h in the FFT periodogram. The detected period of EDA by the CyHMMs is also 24.85 hours. As shown in Table A.3, the results from the three methods reflect that EDA has a more extended period (around 25 hours) than the other three physiological signals. Besides 24 hours, from Fig A.5, the period of 12 h is also significant in the Chi-square periodogram of EDA, heart rate and skin

temperature, and the FFT periodogram of heart rate.

Physiological Data	Algorithm	Detected Periods
BVP	FFT	192.00, 384.00, 128.00, 24.00
	Chi	23.95
	CyHMMs	23.97
EDA	FFT	384.00, 25.60, 24.00, 16.69
	Chi	50.12, 12.00, 24.00
	CyHMMs	24.85
Heart Rate	FFT	24.00, 12.00, 38.40, 11.63
	Chi	23.97, 12.00
	CyHMMs	24.00
Skin Temperature	FFT	384.00, 24.00, 192.00, 128.00
	Chi	23.59, 12.00
	CyHMM	23.83

Table A.3: Periods of BVP, EDA, heart rate, and skin temperature detected by FFT, Chi-square periodogram, and CyHMMs. BVP, heart rate, and skin temperature own 24 h cycle, but the periodicity is not as evident as that of heart rate and skin temperature. The EDA detected by the three algorithms is about one hour longer than that of BVP, heart, and skin temperature.