WiSleep: Smartphone-driven Sleep Population Monitoring with Unsupervised Learning

PRIYANKA MARY MAMMEN, University of Massachusetts Amherst, Amherst, United States
CAMELLIA ZAKARIA, University of Toronto, Toronto, Canada
TERGEL MOLOM-OCHIR, Duke University, Durham, United States
AMEE TRIVEDI, AT&T Inc, Dallas, United States
PRASHANT SHENOY, University of Massachusetts Amherst, Amherst, United States
RAJESH BALAN, Singapore Management University, Singapore, Singapore

With sleep deprivation being a public health concern, sleep monitoring technology, mainly through consumer-grade wearables, has shown value among users to better understand their most fundamental measure of health. Unfortunately, utilizing wearable technology is bound to the conditions of users owning these devices and using them at bedtime every night. While wearables can deliver highly personalized sleep insights to users, they inadvertently affect the ability of sleep monitoring solutions to reach unprivileged sections of society due to added costs and device accessibility. With our primary motivation to promote sleep monitoring for public health use cases at the population scale, we developed *WiSleep*, a sleep monitoring system that infers sleep duration from solely relying on a user's smartphone without requiring a wearable device. Unlike prior efforts that use supervised learning methods and require labeled training data to train sleep models, our method is based on unsupervised learning, which enables easy deployment to new population groups or new regions without a need for labeled data collection and training. Specifically, we employ the smartphone activity of the user, represented by time series of WiFi network event rates, as input data to infer the user's sleep duration (i.e., sleep time and wake time) through an unsupervised Bayesian change point detection ensemble model. Our evaluation shows *WiSleep*'s efficacy in being a low-cost accessible sleep monitoring approach. We present results that yield comparable performance to prior techniques, particularly those requiring new users' labeled data to achieve model personalization. System evaluation from a user study achieved an average of 93.68% accuracy within 59 minutes of sleep time error, 31 minutes of wake time error, and 57 minutes of sleep duration error by utilizing coarse-grained time series data. We demonstrate the application of our technique to predict sleep for 1,000 anonymous users and enable population-scale analytics with low computational overhead.

CCS Concepts: • **Human-centered computing → Ubiquitous and mobile computing**; • **Applied computing → Health informatics**; • **Computing methodologies → Machine learning;**

Additional Key Words and Phrases: Sleep, public health, WiFi, unsupervised learning

This work was supported in part by NSF grants 1722792, 23091241, 2211302, 2211888, 2213636, and 2105494, and US Army grant W911NF-17-2-0196.

Authors' Contact Information: Priyanka Mary Mammen, University of Massachusetts Amherst, Amherst, Massachusetts, United States; e-mail: pmammen@cs.umass.edu; Camellia Zakaria, University of Toronto, Toronto, Ontario, Canada; e-mail: camellia.zakaria@utoronto.ca; Tergel Molom-Ochir, Duke University, Durham, North Carolina, United States; e-mail: tm431@duke.edu; Amee Trivedi, AT&T Inc, Dallas, Texas, United States; e-mail: ameetrivedi@gmail.com; Prashant Shenoy, University of Massachusetts Amherst, Amherst, Massachusetts, United States; e-mail: shenoy@cs.umass.edu; Rajesh Balan, Singapore Management University, Singapore, Singapore, Singapore; e-mail: rajesh@smu.edu.sg.

ACM 2834-5533/2025/02-ART4
https://doi.org/10.1145/3705722

ACM Reference Format:
Priyanka Mary Mammen, Camellia Zakaria, Tergel Molom-Ochir, Amee Trivedi, Prashant Shenoy, and Rajesh Balan. 2025. WiSleep: Smartphone-driven Sleep Population Monitoring with Unsupervised Learning. *ACM J. Comput. Sustain. Soc.* 3, 1, Article 4 (February 2025), 25 pages. https://doi.org/10.1145/3705722

1 Introduction

Sleep is a vital activity that significantly impacts human well-being, productivity, and performance [47]. Prior research has shown that 30% of the adult population does not get enough sleep, with many adults sleeping less than 7 hours per day [18, 31]. Among young adults, particularly college students, sleep hygiene is known to be poor due to variable sleep patterns from academic and social activities. The repercussions of sleep deprivation leading to serious health consequences such as heart disease, stroke, and depression [2, 42] has become a public health burden. Yet, sleep remains an intimate experience; hence, sleep health monitoring requires collecting individual-level user data, and analyzing such data for public health understanding and benefits is highly challenging [42].

While technological advancement offers a broad spectrum of wearable devices [5, 17] and contactless methods using Doppler radar or RF [25, 44] to support personalized sleep health, its continued realization is often impacted by practical user challenges—specifically disruptive device usage or device unavailability from different socioeconomic phenomena [14]. In contrast, as reported by Deloitte [14], the market penetration rate of smartphones is much higher than that of other smart devices, including developing countries. The global smartphone penetration rate is estimated to be around 80% to 82%, implying that most digital users are smartphone owners [14]. As a result, the use of smartphones for monitoring sleep is a viable and more accessible approach, given its broad penetration across the world.

Indeed, prior research focusing on smartphone-based sensing has shown associations between device usage and the users' sleep habits. These examples include analyzing screen activity logs [38], audio from microphone [23], or network activity traffic [26, 37]. Most of these prior approaches are based on machine learning, where a machine learning or a deep learning model is trained using labeled sleep data (i.e., ground truth) collected from users. Such approaches fall in the category of supervised or semi-supervised learning, and use machine learning classifiers to detect sleep and wake states as a proxy of actual sleeping behavior [37]. However, a model trained for one population group (e.g., adults in the United States) may not generalize to a different group (e.g., adolescents in Africa) due to differences in culture and sleep patterns across social groups. As a result, new training data is needed to retrain models prior to deployment in a new region or for a new population group. Since collecting labeled ground truth data at scale is laborious and expensive, we focus on unsupervised learning methods that are based on the concept of priors and do not require training data.

With the overarching goal of building an equity lens to sleep health analytics and supporting a sustainable approach to good health and well-being [51], our research aims to operationalize sleep sensing by asking two key questions: *How can sleep health analytics be available to users who do not own a sleep track or wear the device to bed? How can sleep health models be built without requiring direct observation of new users?* Our overall approach is aligned with the United Nation's sustainable development goal of good health and well-being and bringing these benefits to all sections of society [51].

To address this challenges, this article presents *WiSleep*, a scalable sleep monitoring approach that uses smartphone network activity to infer sleep since such data is available across smartphones whenever users access the Internet. We explore WiFi device connections as an example of smartphone network activity to facilitate device-agnostic capabilities without needing a specific device but a smartphone. Then, we employ an unsupervised learning approach based on Bayesian change point detection to predict sleep periods. The coarse granularity of network activity data naturally limits the prediction capability for precise sleep quality information. However, the singular information on sleep period (duration) remains a useful metric for population-level statistics and can be a precursor to nudge users for individual assessment. The employment of an unsupervised learning technique is necessary to address modeling for new users without ground truth information. In designing, implementing, and evaluating *WiSleep*, our contributions are as follows:

— We evaluate our unsupervised learning approach through a user study involving 20 college students residing in campus dormitories and a private house owner over 1 month. While our approach can handle different population groups, we specifically choose college students for our evaluation, since students are known to have variable sleep patterns, allowing us to test the robustness of our approach. Further, college students, with their variable sleep patterns, often experience sleep deprivation, stress, and mental health issues, making them a particularly important population to study for sleep monitoring. Our results yielded an average accuracy of 93.68%, 87.50% precision, 93.33% recall, and 0.903 F-score within 59, 31, and 57 minutes of sleep, wake time, and sleep duration error.
— We demonstrate the application of our technique to enable sleep population analytics while still rendering personal benefit to individual users. Our analysis on anonymized data of 1,000 on-campus student residents at an aggregated level supports prior findings of students' sleeping patterns over a typical semester. The same platform can provide tailored insights for consented users to understand their sleeping habits by the hour of day and day of the week.

2 Background and Related Work

In this section, we present background and prior efforts related to sleep monitoring. Broadly, sleep monitoring approaches fall into two categories: sleep duration monitoring, which determines how long a user sleeps and when, and sleep quality monitoring, which determines the various stages of sleep and how long the user spends in each stage. Both categories of techniques have been studied using IoT and wearable devices, which are increasingly accepted for everyday use. Similarly, many of the prediction modeling approaches are catered to address individual users' behavior. While our primary focus is sleep duration detection (i.e., detecting the bedtime and wake-up time of a user), we also review methods discussed for measuring sleep quality. Sleep quality encompasses information on sleep stages, such as light sleep, deep sleep, REM (Rapid Eye Movement), and NREM (Non-Rapid Eye Movement). Next, we present background and related work on the sensing modalities used to detect sleep, then focus on various types of prediction algorithms used for this purpose. Table 1 provides an overview of the modality and algorithmic techniques.

2.1 Sensing Modalities for Monitoring Sleep

Sensing sleep for a user requires delivering two functions: the duration and its quality. While both sets of information are equally useful, the modalities and techniques used to derive such information differ significantly. Specifically, sleep quality requires analyzing more fine-grained physiological signals that can more accurately determine one's body movement (e.g., accelerometer) and breathing patterns (e.g., heart rate variability, microphone, blood oxygen saturation) to identify periods of wakefulness during the night and assess their impact on overall sleep quality. Compared to sleep quality, research on sleep duration has shown feasibility with coarse-grained data to determine behaviors of 'going to bed' and 'waking up.'

The modalities for sensing sleep duration have accordingly explored contact-based and contactless methods. Wearable trackers leveraging only accelerometers [32] and mattresses embedded

Table 1. Summary Comparison of Prior Sleep Duration Estimation Approaches

Approach	Scalable	Portable	Labeled Data	Deployment
Motion Sensor [33]	no	no	no	Building
Pressure Sensor [28]	no	yes	yes	Mattress
Phone Activity [13, 22, 23, 38, 46, 58]	yes	yes	yes	Smartphone
Screen Activity [1, 13]	yes	yes	no	Smartphone
Web Activity [3]	yes	yes	yes	Search Engine
WiFi Traffic [26]	yes	yes	yes	WiFi
WiFi Activity [35, 37, 57]	yes	yes	yes	WiFi
WiFi Activity (WiSleep)	yes	yes	**no**	WiFi

Our research particularly contributes to the existing literature by proposing a technique that does not require labeled data from new users for sleep prediction.

with pressure sensors can detect sleep states [28]. Contactless solutions exist to overcome user pushback about wearing devices to sleep by installing sensors in the environment (e.g., wall sensors [33]). By design, contact-based and contactless modalities are appropriate for individual monitoring and are expensive for large-scale monitoring.

In contrast, the ubiquity of smartphones has motivated researchers to exploit the capabilities of phone sensors for sleep monitoring (e.g., analyzing data from microphones [9], cameras [10], and phone activity logs [38, 58]). For example, analyzing screen activities has shown feasibility in inferring sleep and wake [1, 13] due to the strong correlation between (or the lack of) phone activity and the user's sleep. Another work has explored the strong correlation between sleep duration and web activity of users across different times of the day [3]. With users increasingly owning multiple smart devices, the tendency of them switching between devices may not provide a complete representation of users' activities from systems monitoring only their smartphone. Thus, research has also explored monitoring multiple user devices including smart home equipment [57].

2.2 Sleep Prediction Techniques

Regardless of the sensing modality, a variety of machine learning techniques have been developed to infer sleep duration from the sensed data. These techniques can be classified as being supervised or unsupervised. Supervised approaches require large amounts of training data to build sleep detection models. For example, Zakaria et al. [57] developed separate prediction models for each user in the study, which require at least 2 weeks of labeled data with sleep/wake-up estimation errors of less than 40 minutes. Similarly, Mammen et al. [36, 37] adopted a semi-supervised approach, which needs labeled data for developing the baseline model that is further personalized for each user (using unlabeled data from new/unseen users). With the limitations of acquisition of labeled data for model training, current sleep prediction techniques prove to be inadequate as a low-cost sleep monitoring solution. In contrast, unsupervised approaches, such as Bayesian methods, do not need any training data and are easier to deploy at a population scale. For example, Khadiri et al. [29] and Cuttone et al. [13] employed unsupervised Bayesian inference to infer sleep periods using different types of sensors [13, 29]. Although convenient, these works have reported bedtime and wake-time errors in the range of 1 to 2 hours, and they expect users to follow similar bedtime and wake-up time patterns.

As our main goal is to develop sensing mechanisms that can facilitate population-scale prediction, this requirement expects models that can perform accurately for users with varied sleep patterns. Our work seeks to investigate unsupervised learning techniques suitable for users with varied sleep patterns, and demonstrate its application in on-campus student living as a large-scale study.

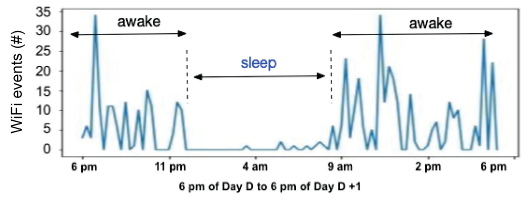


Fig. 1. Smartphone network events over 24 hours, with a low event rate corresponding to sleep.

3 Motivation and Approach

The difficulty and cost of collecting quality labeled data from end users inform our decision to investigate how a single accessible modality and an unsupervised learning technique can be utilized to enable population-scale sleep monitoring.

3.1 Passive WiFi-Sensing Mechanism

As humans grow increasingly reliant on their smartphones [53], much of the common online access, such as video streaming, mobile gaming, and virtual communication, demands low latency and high bandwidth networks that WiFi can offer [48]. Because of this, WiFi is more preferred as a home network solution, running more efficiently in the long run than relying on cellular networks [54]. In this regard, prior works have employed WiFi logs to predict a range of everyday human behavior, such as respiration [30], social interaction, routined activities [24, 27, 56], and sleep monitoring [37]. Using the same modality, our work investigates its use in estimating sleep duration. Unlike prior work that recommends monitoring at least two personal devices (i.e., smartphone, laptop/tablet, smart-TV/speakers), our approach is geared toward minimalist sensing, requiring only WiFi network event rates generated from the user's smartphone as their primary device. Simultaneously, this approach seeks to balance user convenience while requiring minimal user input for accuracy. Although we use a smartphone as the user's primary mobile device, our algorithm can function with any mobile device belonging to the user, such as a laptop or tablet.

We hypothesize that the network activity from a user's phone alone is strongly correlated to the user's awake states and thus is sufficient to infer their sleep periods. Consider the time series example of a user's smartphone network events throughout a 24-hour period to understand why this is feasible. Figure 1 illustrates a user's smartphone network events over a 15-minute interval from Day 1, 6 pm to Day 2, 6 pm. When a user establishes a WiFi connection on their phone for online communication, the device will connect to a nearby **access point (AP)**, generating network events. The device will periodically reassociate to stay connected to the best AP for as long as the user needs the connection, thus triggering a sequence of association and disassociation events. The device eventually falls into a power-saving state when the user stops interacting with it. Periodically, the device 'wakes up' (e.g., every 15 to 30 minutes) and performs a network scan that triggers a reassociation. The fluctuations in network events help us predict the user's activity and state. Similar to techniques relying on a phone's screen activity, we expect that long periods of low network activity are correlated to sleep periods. The main challenge is in determining which period of low network activity should accurately infer a user as (actually) sleeping.

3.2 Unsupervised Learning Technique

The challenge in developing a robust model to support sleep prediction at the population scale is collecting labeled data, which is often labor intensive and expensive. Further, the quality of labeled data depends on accessibility to user groups and their compliance, requiring them to wear

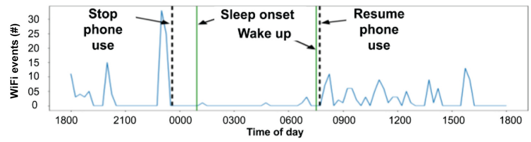


Fig. 2. Potential sensing errors between ceasing/resuming phone activity and sleep/wake onset.

the device and consistently participate in providing ground truth (i.e., recording their sleep habits). Studies have reported decreasing retention rates as a common challenge in large-scale and longitudinal sleep studies [52]. In most cases, these studies include incentivization procedures to encourage continued participation. Data collection over short periods and targeted user groups amount to enough data for studies to pursue their investigation on supervised learning methods, which might not be a viable option in the long run for public health use cases.

In this work, we use an ensemble Bayesian change point detection method where we try to identify the change points—sleep time and wake-up time in the time series data of network events with multiple priors to accommodate different sleeping patterns.

3.3 System Challenges

Any approach that uses a phone's activity as a proxy for a user's activity to infer sleep and wake periods needs to handle many challenges. Both users' behavior and device behaviors can introduce errors. From the user's perspective, it is possible that errors are introduced as a result of the following:

(1) *Sleep onset latency*: A user may not immediately fall asleep upon putting their phones away at night, much more than they would from waking up in the morning (e.g., respond to an alarm).
(2) *Delayed morning smartphone routine*: A user may not check their phone as soon as they wake up.
(3) *Intermittent wake-ups*: A user briefly wakes up during their nocturnal sleep but may not use their phone.

From the device's perspective, it is possible that errors are introduced as a result of the following:

(1) *Ping-pong effects*: The smartphone automatically switches its network connection between nearby APs for the most optimal connection, thus this network activity incorrectly infers user behavior.
(2) *Background activities*: The smartphone performs activities independent of the user, such as installing a software update, as well as receiving emails and messages, incorrectly inferring user behavior.
(3) *Device switch-off/sleep mode*: When a smartphone disconnects from WiFi due to sleep mode for battery savings, battery discharge, or switching to cellular networks, it may incorrectly infer this as an inactive period.

Our approach aims to infer users' sleep duration—and not detect the nuances in sleep characteristics [39, 44] that prior work is set to achieve. For example, as illustrated in Figure 2, ceasing phone activities before sleep time does not immediately translate to sleep onset, as users may take some time to fall asleep. For such reasons, it is difficult to tackle our work as a simple binary classification problem where the longest sequence of low activity periods over a day is determined as the sleeping period.

The rest of the article describes how challenges related to device behaviors are addressed and can support compelling utility at a population scale. We demonstrate through our primary use case, focusing on inferring sleep among the student population and its utility for the student's health and well-being services. A supplementary study on a private home shows how our approach can be applied for personal use.

4 Bayesian Sleep Inference

Our detection mechanism is an ensemble method based on the Bayesian change point model. In what follows, we describe the problem statement to build our model.

Problem Statement. As stated in Section 3, we assume the primary users' device to be smartphones. Consider an enterprise WiFi network deployed in a university campus with M APs and N users. We model each user as being in one of two states: *awake* or *sleeping*. When the user is 'awake,' they can either be mobile (moving from one location to another) or localized at a given area and assumed to use their phone from time to time frequently. In either case, the phone generates AP association and disassociation events logged by the network (we explain this in detail in Section 5.1). With a 24-hour trace of timestamped WiFi events, we use this trace to compute the *rate of network events*; we divide the 24-hour period into time slots and count the number of events in each slot. Let w_t denote the WiFi event rate seen at time t, and let b denote the slot size (we choose a default slot size of b = 15 minutes, yielding 96 slots per day). Given a time series of event rates w_t, our problem is to estimate the sleep onset time, T_{sleep}, and the wake-up time, T_{awake}, for the user.

4.1 Bayesian Change Point Detection

We estimate the sleep and wake-up times from WiFi events based on Bayesian change point detection, which is well established to detect significant changes in time series data and has been widely used for anomaly detection. As illustrated in Figure 1, we must as accurately as possible detect a significant drop in the phone's network activity that occurs at sleep time and a corresponding rise that occurs upon a wake time. Hence, T_{sleep} and T_{awake} are significant change points that we must detect in our time series data w_t, based on Bayesian inference of change points.

We model w_t as a Poisson process (i.e, a time series of event rates in a time slot is Poisson), where λ is the mean of the distribution.

$$P(w) = Poisson(w, \lambda) = \frac{\lambda^w e^{-\lambda}}{w!}$$

Since the mean event rate λ drops at sleep onset time T_{sleep} and rises at wake-up time T_{awake}, therefore λ_{sleep} and λ_{awake} denote the mean event rate when a user is asleep and awake.

$$\lambda = \begin{cases} \lambda_{sleep}, & \text{if } T_{sleep} \leq t < T_{awake} \\ \lambda_{awake}, & \text{otherwise} \end{cases} \tag{1}$$

Since the mean event rate λ_{awake} is high when the user is awake and the event rate λ_{sleep} is low when asleep (see Figure 1), we assume that λ follows a gamma distribution with the following density function.

$$\Gamma(\lambda, a, b) = \frac{1}{\Gamma(a)} b^a \lambda^{a-1} exp(-b\lambda)$$

Given these assumptions, we need to detect two change points T_{sleep} and T_{awake} when the event rate in the time series transitions from λ_{awake} to λ_{sleep} and vice versa. Bayesian change point detection involves finding the posterior distribution of the change points for different values of t and maximizing it to derive the MAP (Maximum A Posterior Estimates). This is done by using a Metropolis-Hastings algorithm [11] to estimate these parameters for each value of t and choosing the t that corresponds to MAP as the change point. As in any Bayesian approach, we need to assign priors to the model parameters (i.e., λ_{sleep}, λ_{awake}, T_{sleep}, T_{awake}) and then use Metropolis sampling to derive the posterior conditional distribution of each parameter from its joint distribution. As noted earlier, the value of t where the distribution is maximized (MAP) represents the change point T_{sleep} (and T_{awake}).

4.2 Ensemble Model for Sleep Inference

The need for our Bayesian approach to be robust to noisy WiFi data and irregular sleep patterns (see Section 3.3) makes it challenging to build a model with strong priors—consequently, models with weak (or non-informative) priors impact model accuracy. Accordingly, we employ an ensemble method comprising three separate models, each with priors suitable for different scenarios, and, finally, apply Bayesian model averaging [19] to derive a combined estimate. The composition of our ensemble model is presented next.

Model 1: Bayesian Model with Location-Based Non-Informative Prior. This model assumes that the sleep periods occur in one or a small subset of locations, such as a dorm room. The location information is inferred directly from the AP placements without localizing the device itself. Priors for a particular day are chosen based on the times spent at these locations. This model is useful for users who have irregular sleep hours but consistent sleep locations. Such location-based priors avoid choosing time periods spent outside the dorm areas for possible sleep periods.

To specify the prior for a specific day, we assume that the mapping of all campus APs to their building locations is known *a priori* and only consider a subset of APs located in the residential dorms. For every user's 24-hour WiFi trace, we determine the longest duration spent in a dorm building (based on network activity observed by the dorm APs). Note, however, that this assumption ignores sleeping behaviors outside the dorm area.

Let $[T_{start}, T_{end}]$ denote the time interval spent in dorm areas, k hours as the minimum sleep duration (e.g., $k = 3$ hours is equivalent to 12 time slots of 15-minute intervals). Since sleep patterns can be irregular, we assume that T_{sleep} and T_{awake} are uniformly distributed within $[T_{start}, T_{end}]$. Hence, the model priors are given as follows.

$$T_{sleep} \sim DiscreteUniform(T_{start}, T_{end} - 12)$$

$$T_{awake} \sim DiscreteUniform(T_{start} + 12, T_{end})$$

The event rate while awake is assumed to be 2.5 events/bin yielding a prior.

$$\lambda_{awake} \sim Gamma(2.5, 1)$$

The event rate while sleeping is assumed to be a low non-zero rate.

$$\lambda_{sleep} \sim Gamma(1, 1)$$

Model 2: Bayesian Model with Normal Prior. This model assumes that the sleep onset and wake-up times are normally distributed (rather than uniformly distributed as in the previous model), thus suited for users with regular sleep and wake-up times.

Let T_{start} and T_{end} denote the start and end times of their daily sleep period. T_{start} and T_{end} are normally distributed with a standard deviation σ. Assume that a student goes to sleep at 12:00 am and wakes up at 8:00 am the next day, with a standard deviation of 3 hours (T_{start} = 12 am, T_{end} = 8 am, σ = 3). The priors for λ_{sleep} and λ_{awake} are the same for all models. Hence, the model priors are given as follows.

$$T_{sleep} \sim Normal(T_{start}, 12)$$

$$T_{awake} \sim Normal(T_{end}, 12)$$

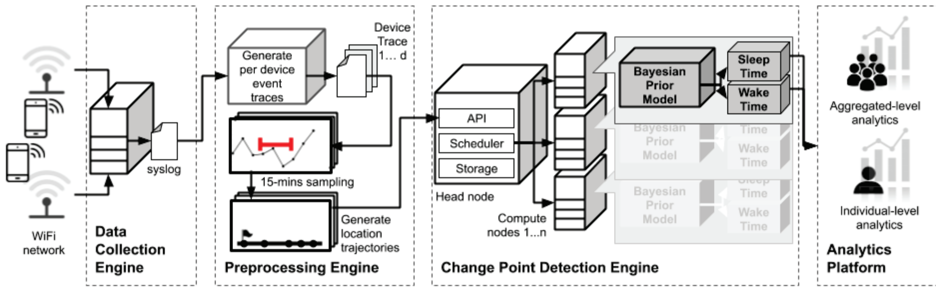


Fig. 3. System components of *WiSleep*.

Model 3: Bayesian Model with Hierarchical Prior. This model is useful when sleep behavior changes based on the day's events, resulting in varying standard deviation.

Let T_{start} and T_{end} denote the start and end times of a sleeping period, normally distributed as per Model 2 (T_{start} = 12 am, T_{end} = 8 am). As sleep behavior varies based on the day's events, T_{sleep} and T_{awake} can be derived by adding hyper-priors α_t, β_t, and τ_t to the normal priors. We set the hyper-priors to a non-informative distribution since we have no strong knowledge about them. The priors for λ_{sleep} and λ_{awake} are the same for all models. Hence, the model priors are given as follows.

$$\begin{aligned}\alpha_t &\sim Exponential(1)\\ \beta_t &\sim Exponential(1)\\ \tau_t &\sim Gamma(\alpha_t, \beta_t)\\ T_{sleep} &\sim Normal(T_{start}, \tau_t)\\ T_{awake} &\sim Normal(T_{end}, \tau_t)\end{aligned}$$

Once all models are utilized for change point detection, these results are averaged using Bayesian model averaging. All models are weighted using a marginal likelihood where the weights are sensitive to the prior distribution. We generate the weights from the posterior distribution of these models using WAIC (Watanabe-Akaike Information Criteria) [55]. WAIC relies on the complete posterior distribution rather than on a single point estimate, making it a more robust approach for generating a combined estimate from the ensemble predictions.

5 *WiSleep* System Overview

We now describe each system component of *WiSleep*, as illustrated in Figure 3. We have built a prototype of our system and deployed it in our university campus and a home setting.

5.1 Data Collection Engine

WiSleep utilizes network events generated by a phone when it connects to a network as a proxy of user activity. While we use WiFi network activity for our experiments, we note that cellular network activity would yield similar results and discuss this extension in Section 9. Moreover, a phone's network events can either be logged on the phone itself (e.g., via a mobile app) or can also be logged by the network (e.g., via logs at APs) since the phone's network behavior is also visible to the network. For convenience, *WiSleep* currently gathers a phone's activity traces using network-level logging at each WiFi AP. However, this can be easily switched to device-level logging on the phone to collect the same information. While leveraging various network event types presents more information of the user, *WiSleep* only uses coarse-grained information. In particular, it uses two types of events: *association* and *disassociation*. To obtain these events, we collect event logs that are generated by all modern WiFi APs (e.g., *syslog*) [6]. A residential setting typically consists of a single AP or a WiFi mesh that logs the connection events of devices, and event log collection is possible in any programmable router. In an enterprise setting such as a campus, a network of APs does the logging. As an example, our campus deployment consists of 5,500 HP/Aruba wireless APs, managed by seven wireless controllers. The syslog data is ultimately sent to a central syslog server for data aggregation of multiple IT systems and network components. As will be explained in Section 6.1, collection of WiFi network logs was with the approval agreement from the university's IT department and the homeowner. Each AP keeps a log comprising a sequence of timestamped events in the following format.

```
hh:mm:ss <controller_name> <process_id> <event_subtype> <MAC_addr> <event_body>
```

While there are many fields to this log, the most relevant to our work are *timestamp*, *controller_name*, *event_subtype*, and *MAC_addr*. A device can be identified through its MAC address (*MAC_addr*). When a device connects and disconnects to an AP, the AP will log an event. It is in this *event_subtype* field that we can distinguish different event types, particularly *association* and *disassociation* events. Specifically, when a user's device connects to WiFi, the connection is established with the nearest WiFi AP from where the user is located, generating *association*-type events. The user's device stays connected to the AP for as long as they utilize the network and remain in that location. When they move to another location, the WiFi connection switches to the next nearest AP where the user is now situated. Accordingly, *disassociation* and *reassociation* events will be generated when the user's device moves out of range and reconnects to the network. Similarly, when a user's device becomes inactive, the AP will log a *disassociation* event. The *controller_name* gives us information on where the user is located based on the nearest AP their device is connected to. It is important to note that throughout the whole time, the user is assumed to maintain the same network connection to the campus WiFi but is only switching APs as they move.

5.2 Preprocessing Engine

Our preprocessing engine takes in the syslog data (with an anonymized MAC address) as input. Note that anonymization is performed on our campus IT department's server before data is copied to our system. The engine proceeds by partitioning event logs to construct per-device event logs of each user's primary device; the primary device is one that makes the largest number of daily AP associations (e.g., over a week). We maintain an up-to-date list of user devices to avoid pulling WiFi events from secondary and/or obsolete devices (e.g., a user may change their smartphone to a new model). Finally, we apply a heuristic to identify devices with high-activity presence in dorm areas as on-campus student residents. The preprocessing engine is written in Python.

5.3 Change Point Detection Engine

Processed per-device event logs are input for our detection engine. It computes WiFi event rates in 15-minute time slots, spanning from 18:00 hours to 17:59 hours the next day. Our model predicts the sleep and wake-up time of users and delivers population-scale and individual-level analytics. We describe our model's performance results in Section 6 and demonstrate our predictive analytics through several case studies in Section 8.

System Performance Metric. Two performance measures are *accuracy* and *timeliness*. As reasoned in Section 4, our engine runs on an ensemble of models based on Bayesian change point detection to yield more acceptable accuracy despite working with weak priors. In Section 6, we present results from comparing the efficacy of *WiSleep* compared to three baseline techniques (i.e., rule-based, normal, and hierarchical priors) and tabulate the prediction accuracies in Table 6.

To achieve timeliness in delivering a population-scale analytics solution, our model utilizes Metropolis-Hastings algorithm [11], which estimates the parameters T_{sleep} and T_{awake} for one user in approximately 4 seconds. We demonstrate in Section 7.4 how *WiSleep* is computationally efficient in producing predictive analytics of 10,000 on-campus student residents under 12 hours. While a single server is adequate to handle the processing needs on our campus, *WiSleep* uses a cluster to scale to larger user populations by parallelizing the analysis of user device traces across servers.[1] In a practical use case for our campus health administrators, *WiSleep* can generate reports of sleep deprivation quickly enough to render pertinent insights.

5.4 Analytics Platform

Results from our unsupervised learning model extend to produce descriptive analytics of sleep patterns among large user groups. Specifically, it profiles anonymized users as regular or irregular sleepers based on their estimated sleep time, wake time, and sleep duration. It generates aggregated reports of these profiles at three time scales—day, week, and month—to provide insight into anonymized profiles with aberrant sleep duration. Section 8 demonstrates several ways our data can be represented and how our findings support prior research on sleep studies, particularly on college students. Further, in Section 9, we discuss how our analytics feature can be operationalized to several end users for public health and personal use while upholding ethical considerations.

6 Experimental Evaluation

We evaluate our model by first, assessing model performance from conducting a study among users living in campus dorms and private housing. Next, we compare model performance with other rule-based and Bayesian techniques.

6.1 Datasets and User Study

Ethical Considerations. This article's data collection and analysis were conducted under safeguards and restrictions approved by our IRB (Institutional Review Board) and Data Usage Agreement with the campus network IT group. All device MAC addresses and authentication information are anonymized using a strong hashing algorithm. User identities were blinded by assigning numeric identifiers. Ground truth was collected within the IRB-approved protocol. It is important to note that our population-scale analysis was performed on aggregate data of anonymous users. Individual analyses were performed on users who had consented to this study.

WiSleep was deployed on our campus and gathered event logs of all connected devices. Our university has more than 31,000 students and close to 14,000 on-campus student residents. With approximately 58,000 detected devices, we anticipate that 14,000 of these devices are applicable for our sleep analyses.

Table 2 summarizes our datasets. Data from our small-scale study was used for model validation. Our small-scale study was conducted in fall 2021 on campus among 20 undergraduates and a single home-user. We precisely identified the participants' hashed MAC addresses by monitoring their WiFi events from a dedicated AP on campus. Each student was given a Fitbit and kept diary logs for ground truth in all three phases. Simultaneously, we collected WiFi events of one homeowner for off-campus private housing validation. His event logs were collected from a home WiFi router. Separately, our case study consists of per-devices event logs of 1,000 students for a given day, demonstrating the types of sleep analytics that *WiSleep* can deliver on a large scale.

[1]Each server is a Dell PowerEdge R430 with a 16-core 2.10-GHz Intel Xeon processor, 64 GB of RAM, 10-gigE network connections, and a local 1-TB disk.

Table 2. Dataset Summary

Study	Gender/Sleep Habit	Dataset
Campus student residents: Fall 2021, 1 month In-home user: 2 weeks	18M 2F (6R, 14IR) 1M (1R)	Identified WiFi network events from smartphone, Fitbit data, diary log
Large-scale weekly analysis	1,000 student residents	Anonymized WiFi network events from smartphone

R denotes regular sleeping habits, and *IR* denotes irregular sleeping habits.

Table 3. Sleep Ground Truth Summary for Campus Student Residents

User Type	Parameter	Median	Mean	Max	Min	Stdev.
Regular	Sleep time	01:45 AM	02:00 AM	10:00 AM	08:00 PM	02:45
	Wake-up time	09:45 AM	10:18 AM	02:00 PM	03:00 AM	02:40
	Duration (hours:mins)	7:15	7:25	10:55	1:00	2:00
Irregular	Sleep time	12:20 AM	01:10 AM	11:00 AM	06:00 PM	03:20
	Wake-up time	09:48 AM	10:08 AM	03:15 PM	10:42 PM	02:40
	Duration (hours:mins)	7:06	6:48	10:55	0:38	1:57

Note: The table values represent aggregated summary for all regular and irregular users.

Table 4. *WiSleep*'s Performance for Different Study Environment

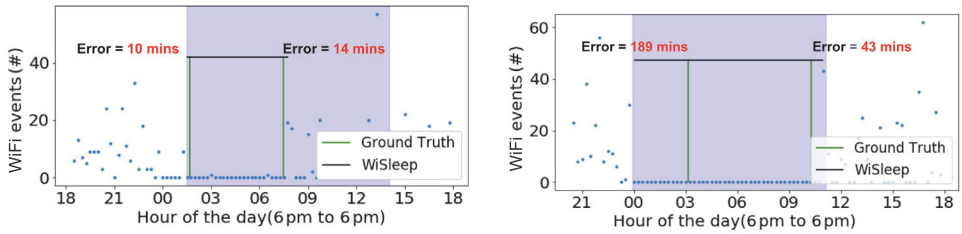
	Accuracy	Precision	Recall	F-score	Sleep, Wake-Up Time Error
Campus Residence	92.7%	87.63%	93.21%	0.90	60, 32 minutes
In-Home	95.8%	87.54%	94.93%	0.91	27, 23 minutes

Table 3 summarizes our participants' sleep logs. We adapted the consistency metric proposed by Rashid et al. [45] to generate a sleep consistency score between 0 and 1 for each user, where 1 denotes the user as having regular sleep patterns throughout the week. We applied a median split to determine the threshold for categorizing users into groups with regular sleep patterns (score 0.75 to 1) and irregular sleep patterns (0 to 0.75). As per Table 2, we identified more users with irregular sleeping patterns than regular ones. The total sleep duration was calculated by subtracting brief awakenings during the user's consolidated sleep period (wake-up time−sleep time). In a healthy adult, the brief awakenings should not be longer than 30 to 45 minutes (considering small measurement errors) [8]. Here, the difference is 42 minutes for regular sleepers (staying within the limits) and 2 hours and 10 minutes for irregular sleepers. This larger discrepancy between irregular sleepers and regular sleepers indicates frequent awakenings and disturbed, delayed sleep onset time, which affects the total time spent asleep versus the time spent in bed.

6.2 Validation Study

We first validate our approach and utilize ground truth data from the user study dataset. We compare our prediction values, T_{sleep} and T_{awake}, with the ground truth Fitbit data and compute four metrics: accuracy, precision, recall, and F-score. Accuracy is the proportion of correct predictions (sleeping or awake periods) relative to all predicted sleeping or awake periods. Precision is the ratio of all correct sleep/awake periods to the total number of predicted sleep/awake periods. F-score indicates the optimal balance that maximizes precision and recall (a score of 1 indicating a perfect predictor).

Table 4 summarizes our final results (after performing error analysis, explained in Section 6.3) with *WiSleep* achieving an average accuracy of 93.68% compared to Fitbit ground truth (+/− 3.90%;



(a) Best case (b) Worst case

Fig. 4. *WiSleep*'s performance for a user, *P*13, living on campus. The shaded area denotes WiFi events in the residential area.

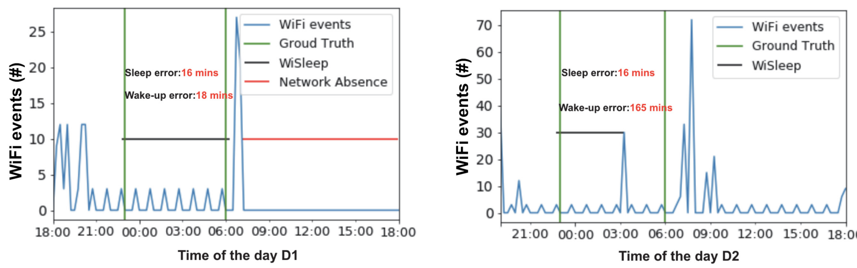
max: 96.00%; min: 81.27%) for predicting users' sleep in a campus residence. In private home use, *WiSleep* yields 95.80% (+/− 2.91%; max: 98.00%; min: 86.00%). As *WiSleep*'s average performance varies between 85% and 95.80% among users, we seek to understand the cases in which our technique performs and breaks. Overall, our system produces error within 59 minutes for estimating sleep time and 31 minutes for wake-up time.

Figure 4 illustrates the WiFi event traces and the predicted/actual sleep/wake-up times of one student resident, *P*13. Figure 4(a) exemplifies the best case prediction, with our model yielding 95.84% accuracy (10 minutes of sleep time error, 15 minutes of wake-up time error, and 6 hours and 35 minutes of sleep duration). In this case, as with most days, the user exhibited near-ideal behavior where sleep onset occurs shortly after ceasing phone activity. Our sleep duration estimation for *P*13 supports prior reporting that the majority of young adults use mobile phones at least 1 hour prior to sleep [20] and at least within 10 minutes after they wake up [43].

In contrast, Figure 4(b) illustrates *P*13's erroneous prediction, with the model performing at 79.25% accuracy (3 hours and 52 minutes of sleep duration). Specifically, *P*13's WiFi event traces did not match his sleep time by more than 189 minutes and wake-up time by 43 minutes, despite the system acquiring no WiFi events between midnight and 11:00 am the next day. As discussed in Section 3.3, deviations in user behaviors can significantly affect system accuracy (e.g., sleep onset latency and delayed use of the smartphone). While the true reasons behind these deviations are undetermined, deviations of user behaviors will pose a limitation. However, these inaccuracies can efficiently be detected as an anomaly.

Private Home Use. To demonstrate the applicability of our system in a private setting, we tested *WiSleep* with one home user and one WiFi AP deployed. In a typical home network setup, a user will have the option to set their mean sleep and wake-up times as part of initializing *WiSleep*. For our user, these times were set to 11:00 pm and 7:00 am, respectively. It is important to note that in this study, we are only tracking a single home user. We discuss in Section 9 how multiple home users can be monitored by *WiSleep*.

Upon running our model for 2 weeks, *WiSleep* successfully yields approximately 95% accuracy (27-minute average sleep time error, 23-minute average wake-up time error). Figure 5 charts the number of WiFi events detected from the user's primary device and, accordingly, his predicted sleep duration and ground truth for two different days, D_1 and D_2. D_1 illustrates an essential aspect of our implementation that does not falsely predict sleep as a result of network absence. On D_1, the user was confirmed to not be present at home between 8:00 am and 6:00 pm. For this reason, the WiFi network captured no network activity, including periodic pings, which would otherwise be recorded had the user (and his primary device) been physically present. In comparison, D_2 represents a day where *WiSleep* underestimates the user's sleep duration. Here, our model incorrectly



(a) Best case (b) Worst case

Fig. 5. *WiSleep*'s performance for a user in a home network.

Table 5. Summary Statistics of Sleep Time Error and Wake-Up Time Error in Minutes

	Ensemble		Hierarchical		Normal		Rule Based	
	T_{sleep}	T_{Wake}	T_{sleep}	T_{wake}	T_{sleep}	T_{Wake}	T_{sleep}	T_{Wake}
Median	43	25	69	58	85	31	202	156
Mean	59	31	109	70	128	80	228	215
Max	189	95	191	100	195	111	250	230
Min	0	0	4	1	1	0	1	1
Stdev.	29	14	51	38	79	43	125	107
Q1, Q3	24, 80	11, 65	26, 92	25, 79	47, 87	12, 61	140, 320	70, 190
UIF, UOF	136, 248	119, 227	158, 290	133, 241	127, 207	110, 208	500, 860	310, 550

inferred the user as waking up from detecting high network activity at 03:15 am, illustrating a deviation in device behavior. We investigate in the next section the breaking point of our model to background activities in Section 7.2. Additionally, a potential solution is employing strong priors in the ensemble model to overcome this device-specific challenge, which would work especially well for users with regular sleeping habits.

6.3 Comparison with State of the Art

Next, we conduct an in-depth analysis by comparing *WiSleep* with prior learning techniques of predicting sleep: a rule-based heuristic and two state-of-the-art Bayesian methods. Our rule-based heuristic first determines a user's residential dorm. It classifies the time (slot) spent in their dorm as active or inactive by checking if the observed WiFi rate is greater than 2 (2 is chosen to ignore the periodic pings). Accordingly, the longest inactivity interval is determined to be the user's sleeping period. The second is a Bayesian approach using normal priors, as proposed by El-Khadiri et al. [15]. The third is also a Bayesian approach using hierarchical priors, as proposed by Cuttone et al. [12].

Table 5 summarizes the descriptive statistics predicting sleep and wake-up times across all methods compared with Fitbit ground truth. Overall, *WiSleep* yields an average error of 59 minutes in sleep time and 31 minutes in wake-up time (*WiSleep*: 93.68%; normal: 85.42%; hierarchical: 87.50%; rule based: 69.79%). The data distribution reveals that a small number of outlier days contribute to a large fraction of the inaccuracies. We calculated the UIF (upper inner fence), the third quartile plus 1.5*IQR, and the UOF (upper outer fence), the third quartile plus 3*IQR, as thresholds for exclusion. Removing outliers resulted in excluding 25 days of prediction among 18 participants, including the examples we presented in Figures 4(b) and 5(b).

Table 6 summarizes our model performance of all techniques compared with Fitbit ground truth. *WiSleep* significantly outperforms the rule-based technique by more than 30% accuracy at

Table 6. *WiSleep*'s Performance Compared against Three Baselines

	Accuracy	Precision	Recall	F-score	Sleep, Wake-Up Time, Duration Error	*p*-Value
WiSleep	93.68%	87.50%	93.33%	0.903	59, 31, 57 minutes	–
Normal [15]	85.42%	75.00%	84.38%	0.794	128, 80, 118 minutes	.01
Hierarchical [12]	87.50%	77.78%	87.50%	0.823	109, 70, 103 minutes	.01
Rule Based	69.79%	54.29%	59.38%	0.563	228, 215, 179 minutes	.001

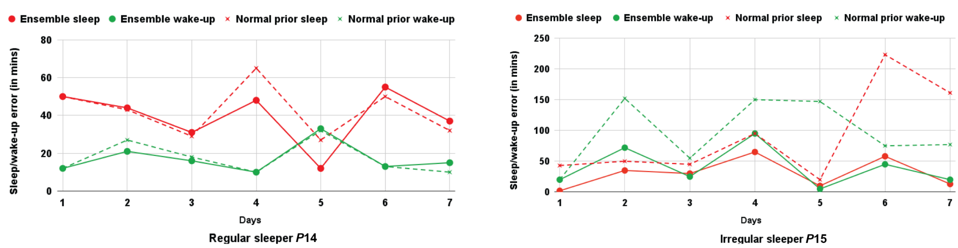


Fig. 6. *WiSleep*'s ensemble versus normal prior for two students with regular sleep (left) and irregular sleep (right) patterns.

$p < .001$. While accuracy improvement is within 6% to 8% compared with other Bayesian techniques, *WiSleep* effectively reduces sleep time error within 60 minutes, wake-up time error within 31 minutes, and sleep duration error within 60 minutes, which is within the acceptable level of clinical acceptance of sleep measurement error [50]. The differences in errors are significantly lesser ($p < .01$) than normal and hierarchical Bayesian. Given that our dataset consists of more irregular sleepers than regulars (see Table 2), the uniform priors, along with location information our ensemble technique employs, will be able to capture different sleeping habits giving weights to appropriate priors.

To illustrate the benefits of our ensemble approach, let us consider two users, $P14$ (a regular sleeper) and $P15$ (an irregular sleeper), as per Figure 6. For $P14$, the ensemble model and normal priors achieve comparable performances overall. In accommodating users with irregular sleep patterns, as per $P15$, the model employing only normal priors would produce larger errors, especially in predicting wake-up times. On days 6 and 7, where sleep time significantly deviates from the population norm, errors from using normal priors were beyond 150 minutes.

7 Practical Considerations

Here, we investigate *WiSleep*'s robustness to device-specific challenges listed in Section 3.3. Additionally, our experiment trials the system to predict an extensive group of users to support large-scale analytics.

7.1 Noisy Data: Ping-Pong Effects

When a stationary device is within the range of connecting to several APs with similar signal strengths, it may connect with one AP. However, the connection can also switch back and forth between different APs, causing a spectrum handoff known as the "ping-pong" effect. The noise from this effect can resemble network activity despite the absence of user interactions. To avoid errors as a result of ping-pong effects, we group APs in an area, such as a dorm floor, and filter out patterns that resemble ping-pongs between nearby APs.

Consider user $P11$, whose phone exhibits significant ping-pong noise, as shown in Figure 7. We observe multiple ping-ponging events between 3:00 am and 4:00 am, but the connection

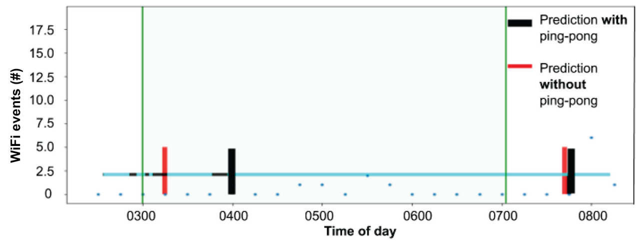


Fig. 7. Ping pong events during $P11$'s sleep period. The green shaded area denotes the ground truth. The cyan horizontal line denotes the primary AP that the user is usually connected to, and black horizontal lines denote other APs in close proximity.

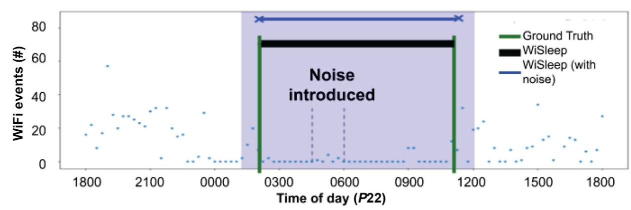


Fig. 8. *WiSleep*'s performance with modest background activities as noise. The shaded area denotes residential area.

remains consistent with the primary AP throughout the rest of the sleep period. Without our heuristic, *WiSleep* would have predicted sleep only after the connection stabilizes at 4:00 am, with a 45-minute delay from the actual sleep time (instead of 3:00 am, closer to the actual value).

7.2 Noise from Background Activities

Background activities such as push notifications and software updates can introduce noise by appearing to have active network usage when a user is asleep. To evaluate the impact of such noise, we carefully introduced background activities in a controlled fashion. Here, we used an Android phone, alternating between long periods of idle, followed by some push notifications, and, finally, a mobile app download from the Play Store. We created a synthetic device trace by inserting noisy traces into an actual device during a nightly sleep period. The synthetic trace, shown in Figure 8, was then subjected to our change point detection method. As shown, the sleep and wake-up times before and after the noise injection are quite similar (≈15 minutes difference in wake-up time).

WiSleep is adaptive to a modest amount of noise from background activities. Since the frequency of push notifications and app updates are typically low, our ensemble method is resilient to noise introduced by their presence. Unfortunately, this also indicates that our technique will produce inaccuracies when high amounts of noise events are present. As discussed in Figure 5(b), a potential solution is employing strong priors in the ensemble model. However, if a user has the habit of streaming movies (which may result in an event peak) at the start of their sleep, the model is likely to predict delayed sleep behavior.

7.3 Impact of Inactive Periods

There can be many device inactivity periods for a user in a day due to sleep mode of the user's device, battery discharge, or the user's physical absence from the network, leading to false positives. *WiSleep* accommodates false positives by picking only the relevant inactive periods using

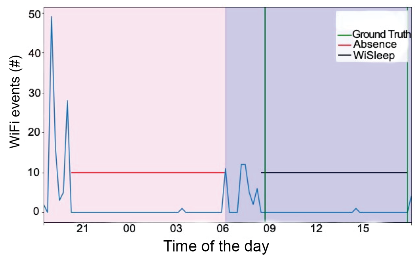


Fig. 9. Impact of inactive periods on *WiSleep*'s performance in non-residential (red shade) and residential (blue shade) areas.

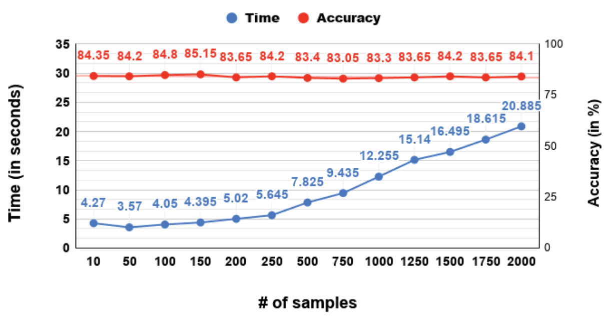


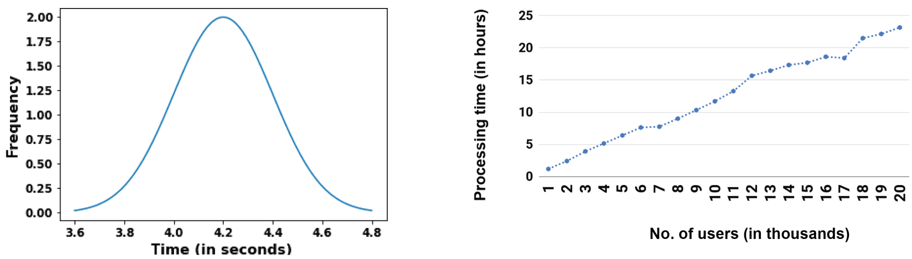
Fig. 10. Accuracy and CPU overhead of change point detection for various sample sizes.

priors for sleep and wake-up times and then considering a user's physical presence in their residential area. For instance, in Figure 9, we observe that a user was inactive at two time periods: first, 8:15 pm to 6:00 am, and second, from 8:15 am to 5:30 pm. Inactivity between 8:15 pm and 6:00 am is typically classified as the eventual sleeping duration, primarily because the user is in the residence. However, this example illustrates a different case—the user is in the residence between 8:15 am and 5:30 pm (this is highlighted in the light blue area). A similar situation can be seen in Figure 5(b), where *WiSleep* can identify network absence, thus avoiding false positives.

7.4 System Scalability

In a real-world implementation of a sleep analytics solution for on-campus student residents, *WiSleep* needs to scale to tens of thousands of users present on campus. Next, we evaluate the scalability of the *WiSleep* system to support a large number of users under accuracy and timeliness constraints. To validate our argument, we examine two factors: (1) the number of samples needed for computation and (2) the CPU cost of the sampling process. First, we determine the number of samples needed for each user to create accurate estimates in the sampling process employed by WiSleep. Generally, the more samples used, the higher the accuracy. However, we must also consider that higher samples will result in higher CPU cost, affecting the results' timeliness.

Figure 10 shows the accuracy and the CPU cost of the computation for two different users obtained by varying the number of samples from 10 to 2,000 over 1 week. We observe that using between 10 and 50 samples yields an accuracy of approximately 85%, which does not significantly change as the sample size increases. Naturally, the more samples used, the higher the CPU cost.



(a) Distribution of time taken per user (b) CPU overheads for 1,000 to 20,000 users

Fig. 11. WiSleep scales up to more than 20k users on a single server.

The results show that a good accuracy–computation tradeoff for *WiSleep* is to use 50 samples producing an accuracy of 85% with a CPU processing cost of approximately 4 seconds per user.

Next, we examine how *WiSleep* scales when processing many users. Figure 11 shows that the CPU time scales linearly with the number of users, and the prediction cycle is completed in 23 hours for 20,000 users, thus showing that a single server is sufficient to handle all on-campus students at our university. Hence, *WiSleep* can generate reports of sleep deprivation of a large number of users quickly enough to render pertinent insights on the same day. One key point is that our system currently uses unoptimized Python libraries for Bayesian inference and does not use any hardware accelerators such as GPUs. With GPU-optimized libraries, the computation time can be reduced further down to less than a second. Additionally, the computation is highly parallelizable and can be scaled near-linearly by using a cluster of servers.

8 *WiSleep* Analytics

We present insights from our large-scale on-campus study of two cases demonstrating how our population-scale aggregate analytics can benefit public health and personal use.

8.1 Population-Scale Aggregate Analytics

Using our large-scale dataset of 1,000 anonymous student users, we conduct an aggregate-level analysis of their sleep behavior for 1 week. Figure 12 plots the average sleep duration of all users by the day. Overall, the results support existing findings that students sleep between 6 and 7 hours, and longer on Sundays [7, 16]. We recognize a slight declining sleep trend at the beginning of the week, before gradual increments later in the week. We expect that the decrease in sleep on weekdays is likely due to students fulfilling various academic demands, whereas Saturday could be attributed to more active extra-curricular activities [16].

To better understand these results, we find that 556 students out of our 1,000 students (>50%) exhibited irregular sleep patterns. This percentage is similar to our small-scale sample as summarized in Table 2. We compare the sleep duration over weekdays and weekends between these sleep profiles, as shown in Figure 13. We observed that, on average, irregular sleepers receive approximately 6.5 hours of sleep, comparable to regular sleepers who get approximately 7 hours of sleep. The aggregated analysis from our large-scale prediction aligns with the ground truth results of our small-scale study (see Table 3).

8.2 Individual-Level Sleep Analytics

Next, we illustrate *WiSleep*'s ability to perform sleep analytics for individual on-campus student users for a semester. We selected two users from our campus user study and retrieved WiFi events

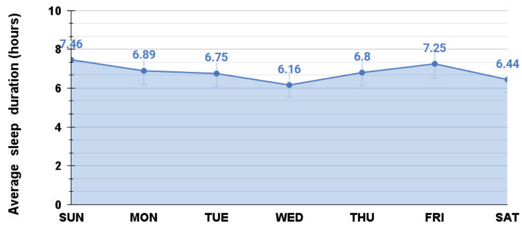


Fig. 12. How do aggregated sleep patterns vary by day of the week? Mean sleep duration predicted for 1,000 users.

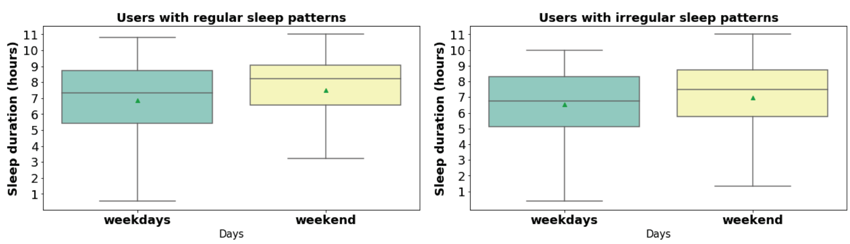


Fig. 13. Box plots comparing the predicted sleep duration difference between users with regular and irregular sleep patterns on weekdays and weekends. Green triangles indicate the mean.

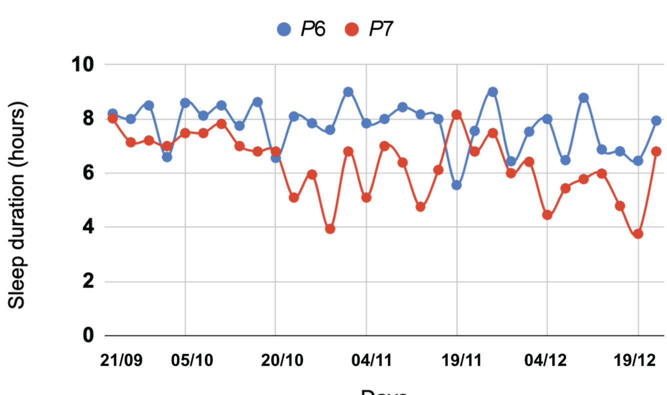
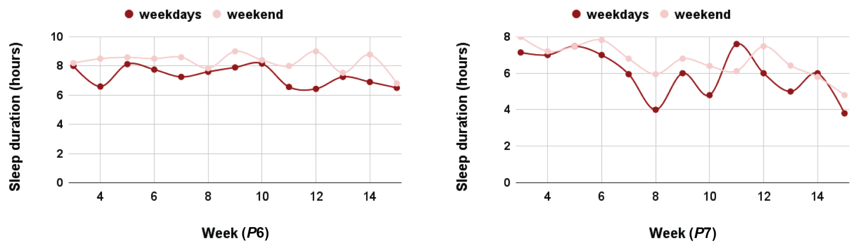


Fig. 14. How do sleep patterns change over a semester? Predicted sleep duration for two participants, *P6* and *P7*.

for approximately 70 days from the start of the semester until the semester end. Note that we intentionally left out the first 3 weeks, as students were more likely to take this time to settle into their student accommodations.

Figure 14 illustrates the predicted sleep duration, averaged every 3 days for two users, *P6* and *P7*. On the whole, both users display sleep inconsistencies throughout the semester. However, *P7*'s sleep patterns seemed fairly consistent at the start of the semester and showed high variability as they transitioned to mid-term week (20/10) until the end of the semester (19/12).

Figure 15 illustrates the sleep regularity pattern for the two users on weekdays and weekends each week over the semester. In general, both users received more sleep on weekends than on



(a) Regular sleeper (b) Irregular sleeper

Fig. 15. Inferred sleep duration for two participants, *P*6 and *P*7, between weekends and weekdays, over the semester.

weekdays: *P*6 weekdays: average sleep = 7 hours and 40 minutes, std = 1 hour and 20 minutes, and weekends: average sleep = 8 hours and 05 minutes, std = 0 hours and 45 minutes; *P*7 weekdays: average sleep = 6 hours and 0 minutes, std = 1 hour and 10 minutes, and weekends: average sleep = 6 hours and 40 minutes, std = 0 hours and 50 minutes. However, on Week 8, corresponding to the mid-terms, we observed *P*7 receiving approximately 4 hours of sleep on the weekdays. This trend persisted throughout Week 10 and Week 13 and reached the lowest in Week 15. On weekends, his sleep pattern showed a consistently declining trend from Week 12 onward.

Indeed, students sleeping less toward the end of the semester is typical. However, users who exhibit a considerable reduction in their sleep throughout extended periods can be a cause for concern. In the following section, we discuss how our system can contribute to opportunities for promoting sleep health.

9 Discussion

We have addressed the challenges of estimating sleep duration and providing analytics for population and individual use. Here, we discuss the implication of *WiSleep* and its limitations.

9.1 Opportunities for Public Health and Personal Well-Being

Indeed, poor sleep hygiene has major health consequences and is a public health issue [2, 21, 42]. *WiSleep* plays a vital role in responding to the call for action to advance sleep disorder problems at the forefront of public health. Among the key strategies promoted, Perry et al. [42] raised the need for improving sleep surveillance among the younger population. In an ideal scenario to strengthen universities' health and well-being services, *WiSleep* can help health professionals decide on critical times to raise community-level awareness on healthy sleep behavior among students. Beyond that, *WiSleep* aggregated analytics can provide "open-access" data sources for public health researchers to study sleep disorders properly. Often, sleep screening is excluded in healthcare screenings compared to those of eating and drinking behaviors [49]. Thus, health professionals do not have enough knowledge of sleep disorders [40].

As stated in Section 6.1, our approach requires access permission on the institutional level—in this case, from the IT department of our university. In practice, the device information is hashed to maintain user anonymity. We had only identified devices of users who consented to the study. Collecting WiFi network event logs does not consist of critical and sensitive information such as browsing content and apps used. This privacy-preserving data modality can allow for sleep analytics without violating data-sharing practices for public health [41].

Deploying interventions remains an option for students who struggle to recover from their lack of sleep, as observed for *P*7 in Section 8.2, and who prefer seeking professional help. To mobilize personal sleep health services, users must consent to be notified through their devices. We imagine that this process will require disclosing their student identification and MAC address of the primary device to progress to a professional health and well-being counselor. Similarly, private homeowners can allow the system to collect WiFi network events logs from their home AP. In a multiple-user home setting, initialization linking users to their devices must be established before monitoring takes place.

9.2 Extensions for Polyphasic Sleep and Sleep Quality

An essential extension is inferring polyphasic sleep. Prior research suggests that people who generally sleep less than 5 hours during the night are more likely to sleep during the day [34] and those who sleep more than 1 hour during the day are more likely to sleep less at night [4]. We could conceivably apply a set of rules to check for secondary sleep on days users are monitored to have slept for less than 5 hours, either before the sleep time ($T_{sleep-1}$) or after the wake-up time ($T_{awake+1}$). Our model could use only a uniform prior (see Section 4.2) to find the longest inactive period at these two times. For instance, in Figure 9, if both inactive periods had occurred in a residential building, the first or second inactive period could be classified as 'secondary sleep.' This warrants further investigation.

Another potential extension to our approach is measuring sleep quality. When implemented at the client side of the network, smartphone activity-based sleep monitoring can be augmented with other phone modalities, such as audio sensing, to detect sleep stages by observing breathing/movement patterns and offer fine-grained sleep monitoring.

9.3 Extensions to Cellular Networks and Client-Side Monitoring

Our WiSleep approach is based on monitoring the network activity of a phone and inferring sleep duration from activity patterns. Several improvements and extensions to the basic approach presented in this work are possible. First, if the network activity can be classified as outgoing network activity and incoming network activity (e.g., the total number of outgoing and incoming packets in an interval), the approach can be enhanced to emphasize outgoing network activity, which is presumably generated by users over incoming network activity, which can also contain background software updates and other non-user data. Second, while we use WiFi network activity for our evaluation, we note that there are regions of the world where users rely on cellular network data for their Internet activities and may not have access to a WiFi network. Fortunately, WiSleep relies on the rate of network events and does not rely specifically on a WiFi network. As long as the rate of events (e.g., packet rates, interface sleep and awake times) can be logged for the cellular network interface, our approach should generalize to cellular networks as well. Finally, network activity can be monitored on the client side or the network side. For convenience, we used network logging as a method to collect a phone's activity logs. However, the same data can be collected on the user's phone via a mobile app and would yield similar results. In fact, for cellular network monitoring, the client-side approach is more viable than a network-wide approach. Finally, both WiFi and cellular data activity can be used together to further enhance the approach. Our focus on college students also allows us to use other devices that are common in student groups, such as laptops or tablets, to infer sleep activities; this may be less feasible, however, for other population groups that largely rely on phones as their main device.

We also note that our design decision to use coarse-grained network events for our model is rooted in maintaining user privacy. By not gathering more fine-grained network activity details, we maintain user privacy. Further, our approach works well even if all network communication is encrypted, since our approach only needs to know when network activity is occurring and is agnostic to the content of the network traffic. Finally, the overhead and energy cost of data collection, whether on the client side or network side, is minimal. On the network side, such coarse-grained events are already logged by the network, whereas on the client side, logging can be implemented efficiently to capture coarse-grained data about network traffic with minimal overheads.

9.4 Limitations

First, our approach assumes that device event data is available on a longitudinal basis for estimating daily sleep duration. However, data for students may be absent from our logs for numerous reasons. Data unavailability will disrupt *WiSleep* from its daily monitoring. Our approach is also not free from periodic maintenance to ensure that user devices are valid (e.g., users did not change their phones) and maintain the current residence (e.g., students may no longer reside in dorms). In addition, a high amount of background noise can sometimes lead to inaccurate predictions, primarily when these noises are attributed to user-specific challenges, such as sleep onset latency or insomnia. Finally, it is essential to emphasize that the goal of our approach is to infer sleep duration of users—and not detect the nuances in sleep characteristics [39, 44] that prior work is set to achieve. The utilization of coarse-grained data inherently limits our ability to determine sleep onset latency, differentiate sleep stages, and detect sleep apnea and insomnia. There is no workaround to this limitation, but our approach remains relevant to achieve the larger goal of complementing population-scale analytics by estimating users' sleep and identifying those with aberrant sleep duration, to which coarse-grained WiFi information is more than adequate.

10 Conclusion

In this article, we presented *WiSleep*, a network-based system to detect sleep periods by passively observing the network activity of a user's phone and provide aggregated and individual-level analytics to accommodate for public and personal use. We presented an ensemble-based Bayesian inference technique to infer sleep from coarse-grained WiFi association and disassociation events. We validated our approach using 20 users living on campus dormitories and one private homeowner in our study. We showed that *WiSleep* outperforms the state-of-the-art methods for users with irregular sleep patterns while yielding comparable accuracy of 93.68% within 59 minutes of sleep time, 31 minutes of wake time, and 57 minutes of duration error. Further, we showed that *WiSleep* can process the data of 20k users on a single commodity machine, allowing it to scale to large campuses with low server requirements. Our large-scale case study revealed several interesting insights for population-scale and individual sleep analytics. As future work, we plan to combine our sleep inference model with stress detection methods to develop a complete student well-being service.

References

[1] Saeed Abdullah, Mark Matthews, Elizabeth L. Murnane, Geri Gay, and Tanzeem Choudhury. 2014. 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.*

[2] Bruce M. Altevogt and Harvey R. Colten. 2006. *Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem.* National Academies Press.

[3] Tim Althoff, Eric Horvitz, Ryen W. White, and Jamie Zeitzer. 2017. Harnessing the web for population-scale physiological sensing: A case study of sleep and performance. In *Proceedings of the 26th International Conference on World Wide Web.* 113–122.

[4] Sonia Ancoli-Israel and Jennifer L. Martin. 2006. Insomnia and daytime napping in older adults. *Journal of Clinical Sleep Medicine* 2, 3 (2006), 333–342.

[5] Apple. 2020. Home Page. Retrieved July 8, 2020 from https://www.apple.com

[6] Aruba Networks. 2016. ArubaOS 6.5.x Syslog Messages, Reference Guide. Retrieved December 13, 2024 from https://www.hpe.com/psnow/doc/c05321932

[7] Walter C. Buboltz Jr., Franklin Brown, and Barlow Soper. 2001. Sleep habits and patterns of college students: A preliminary study. *Journal of American College Health* 50, 3 (2001), 131–135.

[8] Mary A. Carskadon and William C. Dement. 2005. Normal human sleep: An overview. In *Principles and Practice of Sleep Medicine* (4th ed.), Meir H. Kryger, Thomas Roth, and William Dement (Eds.). Saunders, 13–23.

[9] Xiangmao Chang, Cheng Peng, Guoliang Xing, Tian Hao, and Gang Zhou. 2020. iSleep: A smartphone system for unobtrusive sleep quality monitoring. *ACM Transactions on Sensor Networks* 16, 3 (2020), 1–32.

[10] Zhenyu Chen, Mu Lin, Fanglin Chen, Nicholas D. Lane, Giuseppe Cardone, Rui Wang, Tianxing Li, Yiqiang Chen, Tanzeem Choudhury, and Andrew T. Campbell. 2013. Unobtrusive sleep monitoring using smartphones. In *Proceedings of the 2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops*. IEEE, 145–152.

[11] Siddhartha Chib and Edward Greenberg. 1995. Understanding the Metropolis-Hastings algorithm. *American Statistician* 49, 4 (1995), 327–335.

[12] Andrea Cuttone, Per Bækgaard, Vedran Sekara, Håkan Jonsson, Jakob Eg Larsen, and Sune Lehmann. 2017. SensibleSleep: A Bayesian model for learning sleep patterns from smartphone events. *PLoS One* 12, 1 (2017), e0169901.

[13] Andrea Cuttone, Per Bakgaard, Vedran Sekara, Håkan Jonsson, Jakob Eg Larsen, and Sune Lehmann. 2017. SensibleSleep: A Bayesian model for learning sleep patterns from smartphone events. *PLoS One* 12, 1 (2017), e0169901.

[14] Deloitte. 2017. Global mobile consumer trends, 2nd edition: Mobile continues its global reach into all aspects of consumers' lives. *Deloitte Report*. Retrieved December 13, 2024 from https://www2.deloitte.com/content/dam/Deloitte/us/Documents/technology-media-telecommunications/us-global-mobile-consumer-survey-second-edition.pdf

[15] Yassine El-Khadiri, Gabriel Corona, Cédric Rose, and François Charpillet. 2018. Sleep activity recognition using binary motion sensors. In *Proceedings of the 2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI '18)*. IEEE, 265–269.

[16] Andrea K. Finlay, Nilam Ram, Jennifer L. Maggs, and Linda L. Caldwell. 2012. Leisure activities, the social weekend, and alcohol use: Evidence from a daily study of first-year college students. *Journal of Studies on Alcohol and Drugs* 73, 2 (2012), 250–259.

[17] Fitbit Inc. 2020. Home Page. Retrieved December 13, 2024 from https://www.fitbit.com

[18] Centers for Disease Control and Prevention (CDC). 2009. Perceived insufficient rest or sleep among adults—United States, 2008. *MMWR: Morbidity and Mortality Weekly Report* 58, 42 (2009), 1175.

[19] Tiago M. Fragoso, Wesley Bertoli, and Francisco Louzada. 2017. Bayesian model averaging: A systematic review and conceptual classification. *International Statistical Review* 86, 1 (December 2017), 1–28.

[20] Michael Gradisar, Amy R. Wolfson, Allison G. Harvey, Lauren Hale, Russell Rosenberg, and Charles A. Czeisler. 2013. The sleep and technology use of Americans: Findings from the National Sleep Foundation's 2011 Sleep in America poll. *Journal of Clinical Sleep Medicine* 9, 12 (2013), 1291–1299.

[21] Adolescent Sleep Working Group. 2014. School start times for adolescents. *Pediatrics* 134, 3 (2014), 642–649.

[22] Weixi Gu, Longfei Shangguan, Zheng Yang, and Yunhao Liu. 2015. Sleep Hunter: Towards fine grained sleep stage tracking with smartphones. *IEEE Transactions on Mobile Computing* 15, 6 (2015), 1514–1527.

[23] Tian Hao, Guoliang Xing, and Gang Zhou. 2013. iSleep: Unobtrusive sleep quality monitoring using smartphones. In *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems*. 1–14.

[24] Hande Hong, Chengwen Luo, and Mun Choon Chan. 2016. SocialProbe: Understanding social interaction through passive WiFi monitoring. In *Proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services*. 94–103.

[25] Chen-Yu Hsu, Aayush Ahuja, Shichao Yue, Rumen Hristov, Zachary Kabelac, and Dina Katabi. 2017. Zero-effort in-home sleep and insomnia monitoring using radio signals. In *Proceedings of the ACM Interact. Mob. Wearable Ubiquitous Technology*.

[26] Dheryta Jaisinghani and Nishtha Phutela. 2023. Packets-to-prediction: An unobtrusive mechanism for identifying coarse-grained sleep patterns with WiFi MAC layer traffic. *Sensors* 23, 14 (2023), 6631.

[27] Eftychia Kalogianni, R. Sileryte, Marco Lam, Kaixuan Zhou, Martijn Van der Ham, S. Van der Spek, and E. Verbree. 2015. Passive WiFi monitoring of the rhythm of the campus. In *Proceedings of the 18th AGILE International Conference on Geographic Information Science*. 9–14.

[28] Lih-Jen Kau, Mao-Yin Wang, and Houcheng Zhou. 2023. Pressure-sensor-based sleep status and quality evaluation system. *IEEE Sensors Journal* 23, 9 (2023), 9739–9754.

[29] Yassine El Khadiri, Gabriel Corona, Ceedric Rose, and Francois Charpillet. 2018. Sleep activity recognition using binary motion sensors. In *Proceedings of the 30th IEEE Conference on Tools with Artificial Intelligence*.

[30] Usman Mahmood Khan, Zain Kabir, Syed Ali Hassan, and Syed Hassan Ahmed. 2017. A deep learning framework using passive WiFi sensing for respiration monitoring. In *Proceedings of the 2017 IEEE Global Communications Conference (GLOBECOM '17)*. IEEE, 1–6.

[31] Patrick M. Krueger and Elliot M. Friedman. 2009. Sleep duration in the United States: A cross-sectional population-based study. *American Journal of Epidemiology* 169, 9 (2009), 1052–1063.

[32] Xinyue Li, Yunting Zhang, Fan Jiang, and Hongyu Zhao. 2020. A novel machine learning unsupervised algorithm for sleep/wake identification using actigraphy. *Chronobiology International* 37, 7 (2020), 1002–1015.

[33] Feng Lin, Yan Zhuang, Chen Song, Aosen Wang, Yiran Li, Changzhan Gu, Changzhi Li, and Wenyao Xu. 2016. SleepSense: A noncontact and cost-effective sleep monitoring system. *IEEE Transactions on Biomedical Circuits and Systems* 11, 1 (2016), 189–202.

[34] Xianchen Liu, Makoto Uchiyama, Keiko Kim, Masako Okawa, Kayo Shibui, Yoshihisa Kudo, Yuriko Doi, Masumi Minowa, and Ryuji Ogihara. 2000. Sleep loss and daytime sleepiness in the general adult population of Japan. *Psychiatry Research* 93, 1 (2000), 1–11.

[35] Priyanka Mary Mammen and Prashant Shenoy. 2022. Are you asleep when your phone is asleep? Semi-supervised methods to infer sleep from smart devices. In *Proceedings of the 36th Conference on Neural Information Processing Systems (NeurIPS '22): Workshop on Learning from Time Series for Health*. 1–5.

[36] Priyanka Mary Mammen, Camellia Zakaria, and Prashant Shenoy. 2023. Personalized sleep monitoring using smartphones and semi-supervised learning. In *Proceedings of the International Conference on Pervasive Computing Technologies for Healthcare*. 322–338.

[37] Priyanka Mary Mammen, Camellia Zakaria, and Prashant Shenoy. 2023. SleepLess: Personalized sleep monitoring using smartphones and semi-supervised learning. *CSI Transactions on ICT* 11, 4 (2023), 203–219.

[38] Jun-Ki Min, Afsaneh Doryab, Jason Wiese, Shahriyar Amini, John Zimmerman, and Jason I. Hong. 2014. Toss 'N' Turn: Smartphone as sleep and sleep quality detector. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 477–486.

[39] Anh Nguyen, Raghda Alqurashi, Zohreh Raghebi, Farnoush Banaei Kashani, Ann C. Halbower, and Tam Vu. 2016. A lightweight and inexpensive in-ear sensing system for automatic whole-night sleep stage monitoring. In *Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems (SenSys '16)*. ACM, 230–244.

[40] Klara K. Papp, Carolyn E. Penrod, and Kingman P. Strohl. 2002. Knowledge and attitudes of primary care physicians toward sleep and sleep disorders. *Sleep and Breathing* 6, 3 (2002), 103–109.

[41] Michael Parker and Susan Bull. 2015. Sharing public health research data: Toward the development of ethical data-sharing practice in low- and middle-income settings. *Journal of Empirical Research on Human Research Ethics* 10, 3 (2015), 217–224.

[42] Geraldine S. Perry, Susheel P. Patil, and Letitia R. Presley-Cantrell. 2013. Raising awareness of sleep as a healthy behavior. *Preventing Chronic Disease* 10 (2013), E133.

[43] Tyler Schmall. 2018. Americans spend half their lives in front of screens. *New York Post*. Retrieved August 13, 2018 from https://nypost.com/2018/08/13/americans-spend-half-their-lives-in-front-of-screens/

[44] Tauhidur Rahman, Alexander T. Adams, Ruth Vinisha Ravichandran, Mi Zhang, Shwetak N. Patel, Julie A. Kientz, and Tanzeem Choudhury. 2015. DoppleSleep: A contactless unobtrusive sleep sensing system using short-range Doppler radar. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 39–50.

[45] Haroon Rashid, Pushpendra Singh, and Krithi Ramamritham. 2017. Revisiting selection of residential consumers for demand response programs. In *Proceedings of the 4th ACM International Conference on Systems for Energy-Efficient Built Environments*. 1–4.

[46] Yanzhi Ren, Chen Wang, Jie Yang, and Yingying Chen. 2015. Fine-grained sleep monitoring: Hearing your breathing with smartphones. In *Proceedings of the 2015 IEEE Conference on Computer Communications (INFOCOM '15)*. IEEE, 1194–1202.

[47] Mark R. Rosekind, Kevin B. Gregory, Melissa M. Mallis, Summer L. Brandt, Brian Seal, and Debra Lerner. 2010. The cost of poor sleep: Workplace productivity loss and associated costs. *Journal of Occupational and Environmental Medicine* 52, 1 (2010), 91–98.

[48] Joel Sommers and Paul Barford. 2012. Cell vs. WiFi: On the performance of metro area mobile connections. In *Proceedings of the 2012 Internet Measurement Conference*. 301–314.

[49] Adam J. Sorscher. 2008. How is your sleep: A neglected topic for health care screening. *Journal of the American Board of Family Medicine* 21, 2 (2008), 141–148.

[50] Nicola Straiton, Muaddi Alharbi, Adrian Bauman, Lis Neubeck, Janice Gullick, Ravinay Bhindi, and Robyn Gallagher. 2018. The validity and reliability of consumer-grade activity trackers in older, community-dwelling adults: A systematic review. *Maturitas* 112 (2018), 85–93.

[51] United Nations. n.d. United Nations Sustainable Development. Retrieved December 13, 2024 from https://www.un.org/sustainabledevelopment/

[52] Jonas Van Der Donckt, Nicolas Vandenbussche, Jeroen Van Der Donckt, Stephanie Chen, Marija Stojchevska, Mathias De Brouwer, Bram Steenwinckel, Koen Paemeleire, Femke Ongenae, and Sofie Van Hoecke. 2024. Addressing data quality challenges in observational ambulatory studies: Analysis, methodologies and practical solutions for wrist-worn wearable monitoring. *arXiv preprint arXiv:2401.13518* (2024).

[53] Alexander J. A. M. Van Deursen, Colin L. Bolle, Sabrina M. Hegner, and Piet A. M. Kommers. 2015. Modeling habitual and addictive smartphone behavior: The role of smartphone usage types, emotional intelligence, social stress, self-regulation, age, and gender. *Computers in Human Behavior* 45 (2015), 411–420.

[54] Shweta Ware, Chaoqun Yue, Reynaldo Morillo, Jin Lu, Chao Shang, Jayesh Kamath, Athanasios Bamis, Jinbo Bi, Alexander Russell, and Bing Wang. 2018. Large-scale automatic depression screening using meta-data from WiFi infrastructure. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 4 (2018), 1–27.

[55] Sumio Watanabe. 2013. A widely applicable Bayesian information criterion. *Journal of Machine Learning Research* 14 (2013), 867–897.

[56] Camellia Zakaria, Rajesh Balan, and Youngki Lee. 2019. StressMon: Scalable detection of perceived stress and depression using passive sensing of changes in work routines and group interactions. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–29.

[57] Camellia Zakaria, Gizem Yilmaz, Priyanka Mary Mammen, Michael Chee, Prashant Shenoy, and Rajesh Balan. 2023. SleepMore: Inferring sleep duration at scale via multi-device WiFi sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 4 (2023), 1–32.

[58] Chen Zhenyu, Nicholas Lane, Guiseppe Cardone, Mu Lin, Tanzeem Choudhury, and Andrew Campbell. 2013. Unobtrusive sleep monitoring using smartphones. In *Proceedings of the 2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops*.

Received 23 April 2024; revised 14 September 2024; accepted 13 October 2024