Personalized Sleep Monitoring Using Smartphones and Semi-supervised Learning

Priyanka Mary Mammen[1(✉)], Camellia Zakaria[2], and Prashant Shenoy[1]

[1] University of Massachusetts Amherst, Amherst, USA
{pmammen,shenoy}@cs.umass.edu

[2] University of Toronto, Toronto, Canada
camellia.zakaria@utoronto.ca

Abstract. Sleep is a critical aspect of an individual's physical and mental well-being. Hence a large body of sleep monitoring solutions has been gaining popularity, including data-driven AI techniques with mHealth adaptations of wearable, smartphone, and contactless-sensing modalities. Regardless, proposed solutions by prior works, by and large, require gathering sufficient ground truth data to develop personalize and highly accurate sleep prediction models. This requirement inherently presents a challenge of such models underperforming when inferring sleep on new users without labeled data. However, unlabeled data is often more abundantly available in a real-world application than gathering labeled data. In this paper, we propose *SleepLess*, which uses semi-supervised learning over unlabeled data sensed from the user's smartphone network activity to develop personalized models and detect their sleep duration for the night. Specifically, it uses a pre-trained model on an existing set of users to produce pseudo labels for unlabeled data of a new user and achieves personalization by fine-tuning over selectively picking the pseudo labels. Our IRB-approved user study among 23 users found *SleepLess* model yielding around 96% accuracy, between 12–27 min of sleep time error and 18–25 min of wake time error. Comparison against other approaches that sought to predict with fewer labeled data found *SleepLess*, similarly yielding best performance. Our study demonstrates the feasibility of achieving personalization in sleep prediction models by utilizing unlabeled data extracted from network activity of users' smartphones, through the application of a semi-supervised transfer learning approach.

Keywords: Semi-supervised Learning · Time Series data · Sleep · Passive Sensing

1 Introduction

Sleep is an essential daily human activity that significantly affects a person's health and well-being. Despite its importance, sleep disorder is common among

D. Salvi et al. (Eds.): PH 2023, LNICST 572, pp. 322–338, 2024.
https://doi.org/10.1007/978-3-031-59717-6_22

adults, with prior studies reporting 20–40% adults suffering from a form of sleep disorder [5]. Sleep deprivation is a widespread problem, with a third of the population getting less sleep than the recommended 8 h of regular sleep [1]. Since poor sleep hygiene can influence various health problems, sleep monitoring has become a critical technology enabler for researchers and clinicians to understand daily sleep habits better and identify poor sleep health.

Wearable sleep trackers such as FitBit [10] and Oura ring [39] have become popular for users to keep track of their daily sleep in recent years. Although they are simple to use, these contact-based methods may be less favorable among users who prefer not to wear a device during their sleep time [2]. To overcome this challenge, researchers have responded by developing several contactless solutions. For example, radar-based approaches [2], use radio frequency signals that bounce off users to monitor their breathing and sleep. This technology, in particular, is adopted by smart speakers such as Google Nest [6] and Amazon Alexa [7] for contactless sleep tracking using a built-in radar [35]. While wearable and smart speakers can monitor sleep duration and quality, smartphones are more ubiquitous. Researchers have attempted to leverage the ubiquity of smartphones as an inexpensive means of tracking users' sleep. The primary approach of such solutions is based on indirect sensing, where passive observations of smartphone activities are used to infer a user's sleep duration. An early work [3], which utilizes smartphone screen activity as a proxy of their awake states, correspondingly estimating sleep based on users' inactivity. More recent work has successfully generalized this notion to utilizing network activity generated by smartphones and smart devices where long periods of inactivity were used to detect sleep periods [46],?. A general drawback of these solutions is the fundamental need to collect labeled ground truth data from users for training prediction models that will accurately infer their sleep. Due to a myriad of user-related issues, such as inaccurate data logging, missing data, and eventually, user attrition, conducting long-term user studies to specifically collect large amounts of ground truth data is challenging [8]. Consequently, many research efforts to study sleep have been limited to a small sample population of tens of users.

Designing models that generalize to a larger population using a small sample size and a small amount of labeled ground truth data is challenging—especially since sleep patterns can vary from one user to another. In contrast to labeled data that is difficult to collect via user studies, unlabeled data of a phone's network or screen activity is significantly easier to collect via automated apps; neither of these data sources require user involvement during data collection. Similarly, WiFi networks routinely logs client activities which can be used to infer a phone's network activity over time [9]. *The convenient availability of unlabeled data together coupled with the challenges gathering labeled data motivates the need to develop semi-supervised learning (SSL) methods* that can monitor a user's sleep using passive observations of the phone's network activity. Further, such SSL methods should enable personalization of the learnt model in order to handle each user's sleep patterns.

In this paper, we present *SleepLess*, a system that uses semi-supervised learning over unlabeled phone data to develop personalized models for detecting a user's daily sleep patterns. In designing and evaluating *SleepLess*, we make the following contributions:

1. We propose a semi-supervised training pipeline to enable personalized sleep duration estimation in users from the network activity of their mobile-phones. We use a teacher-student framework to utilize a pre-trained sleep prediction model and a few weeks of unlabeled data from the end-user.
2. We implement a complete prototype of a semi-supervised learning pipeline to demonstrate the efficacy of our approach. We conduct a user study on a campus consisting of 20 users. Further, we present a case study to demonstrate the generalizability of the approach in residential settings.
3. The model validations show that our approach achieves around 96% accuracy, between 12–27 min of sleep time error and 18–25 min of wake time error.

2 Background

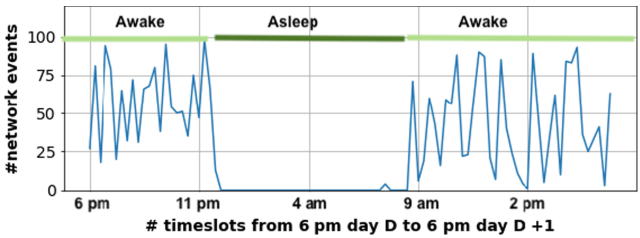

Fig. 1. Time-series data of network activity of mobile phones representing sleep and wake-up periods.

The main premise behind inferring a user's sleep through their smartphone activity is that users utilize their smartphone throughout the day, generating screen or network activity as soon as they are awake. Conversely, the lack of phone-generated activities are more likely to occur when users are asleep. This preliminary insight was leveraged by many prior work to demonstrate the feasibility of inferring sleep periods from approximating user activity through the smartphone alone [3,46]. The same concept can be applied to a phone's network activity as shown in Fig. 1. The figure shows a phone's WiFi event rate every 15 min over a day and clearly shows a low activity period during the night corresponding to the user's nocturnal sleep period.

3 Related Work

Polysomnography is the gold standard for sleep monitoring [36]. However, conducting sleep studies outside clinical settings would require less obtrusive monitoring techniques for long-term sleep monitoring. As an alternative, researchers explored using heart rate sensors and motion sensors, found in commercially available sleep trackers such as Fitbit and Apple watch [40], for detecting human activity [41]. Although such wearables are very convenient to use, there are cases where users may not always wear them to bed. Instead, contactless solutions such as doppler radar or RF signals are deployed as standalone devices to monitor breathing patterns [42,43] and predict sleep [44,45]. Other sensing techniques look into smartphone based sensing techniques for sleep tracking. These efforts utilize an array of phone sensors such as accelerometer [13,22], light sensor [13,22], microphone [13,22–24], proximity sensor [22], and WiFi network activity rates [46]. These works above have proven rich (unlabeled) data we can use to predict sleep. However, collecting labeled ground truth of users continue to be a big challenge [4]. Most of the prior sleep prediction techniques [13,22] work are based on supervised learning approaches, whereby models require large amounts of training data to build accurate prediction models . Several works developed separate models for each user in the study which require at least 2 weeks of labelled data with sleep/wake-up estimation errors less than 40 min [46]. On the other hand, unsupervised methods have been explored specifically to do away with requiring training data. For example, Cuttone et al. develop a Bayesian approach to infer bed time and wake-up time from smart-phone screen events. Although convenient, these works have reported bed time and wake-time errors in the range of 1–2 h and they don't offer any form of personalization. These results collectively informed our decision to explore semi-supervised learning (SSL) approaches, where we can leverage unlabeled data to improve the accuracy of the prediction models.

Our investigation on semi-supervised learning approaches began via Self Training, followed by other standard techniques such as Co-Training, Auto Encoder, Data Generative, and Adversarial Training. These approaches have been adopted by prior work to solve sleep stage classification [15,17,50,51]. In understanding self-training, Zhang et al. reported error accumulation as a potential problem [52]. In contrast, other works have reported lesser error accumulation with co-training [53,54], including successful detection of everyday human behavior such as walking, running, and climbing stairs [55]. In fact, much work in human activity recognition has utilized adversarial learning [56] and autoencoder [29,57] to develop a generalizable and robust classification model for everyday human behavior.

Similar to these works, we aim to predict a person's sleep duration every night as an everyday human activity. However, these works rely on fine-grained time-series data sources such as EEG and actigraphy data, which are more likely to offer data completeness. In contrast, our work aims to develop a prediction model that can leverage unlabeled data, which is also suitable for coarse-grained data.

4 *SleepLess* Design

This section presents the design of our system, *SleepLess*, beginning with its problem formulation.

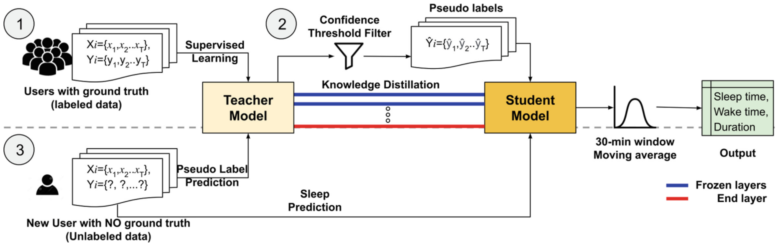

Fig. 2. *SleepLess*'s approach is a three-step process of training a Teacher model, generating high-quality pseudo labels from unlabeled data, and personalizing a Student model for a new user.

4.1 Problem Statement

The goal of our work is to develop personalized sleep detection models for each user based on unlabeled activity data from their smartphones. Our work assumes that a small amount of labeled data is available from a small group of users, which can be used for initial training. Additionally, we assume that only unlabeled data of new users is available, but their model needs to be personalized. We seek to design a semi-supervised learning approach to personalize an existing model, pre-trained on other sets of users, solely using unlabeled smartphone data for the new user. For the purpose of this paper, we consider utilizing coarse-grained WiFi activity data generated by users' smartphones as a measure of their phone activity[1]. Formally, we model this problem as a multivariate time series classification problem:

Users with Labeled Data: Consider $X^i = \{x_1, x_2 ....x_T\}$, which represents the multivariate time series of phone activity features for user, i, where $x_j = \{f_1, f_2 ....f_n\}$ is a vector of phone activity features at time, j. The features, $f_1, f_2 ....f_n$, in our case, are WiFi network activity features generated by user i's smartphone; for example, the number of observed WiFi events, number of WiFi access points connected by the phone. Collectively, these features represent the level of a phone's network activity. We assume time is discretized into fixed length intervals (i.e., 15-min) and these activity features are computed for each interval.

We assume a small group of users whose ground truth sleep information is available. This information includes the user's sleep duration, sleep time, T_{sleep}, and wake time T_{wake}. The ground truth yields a labeled time series for each user, i, where $Y^i = \{y_1, y_2 ....y_T\}$ and the label for each interval j is $y_j \in \{0, 1\}$. A label of 1 denotes the user as asleep, conversely, 0 denotes the user as being awake. All intervals between T_{sleep} and T_{wake} gets a label of 1.

[1] Our approach can be applied to other types of phone activity data such as screen activity. Here, phone activity data is represented by network activity rate.

New Users with Unlabeled Data: We assume a much larger group of users, whose phone activity data X^i is available but no labeled ground truth Y^i is known. In this case, the time series X^i simply represents unlabeled activity data for the user.

Henceforth, the problem is to train an (initially) supervised model on users with labeled data and personalize this model for each new user with only using their unlabeled data.

4.2 *SleepLess* Approach

Our approach to addressing the above problem involves three key steps, depicted in Fig. 2.

Step 1: Train a Teacher Model. *SleepLess* first users the set of users with ground truth data to train an initial CNN-based deep learning model. We refer to this initial model as the teacher model, $Model_{Teacher}$. $Model_{Teacher}$, discussed further in Sect. 5.3, uses a cross-entropy loss function defined as:

$$H_p(q) = -\frac{1}{N}\sum_{i=1}^{N} y_i.log(p(y_i)) + (1 - y_i).log(1 - p(y_i)) \quad (1)$$

where y is the sleep or awake label.

In solving for a binary classification problem, prior work has reported sigmoid value as a confidence metric reliable despite overconfident predictions arising from unseen classes [60]. However, due to the imbalanced nature of the dataset where there are more awake labels than sleep, the model may be biased towards predicting more prevalent label (i.e., awake). To address this issue, we calibrate our models by leveraging a validation set to improve the accuracy of our prediction probabilities.

Note that $Model_{Teacher}$ is trained to perform a binary classification task by taking the phones activity features X_j and predict whether the user is awake or asleep. The longest sequence of sleep labels over the course of each 24-h represent the sleep period for that day.

Step 2: Obtain Pseudo Labels from the Teacher Model. Given the teacher model, $Model_{Teacher}$, *SleepLess* then considers a new user, k, whose phone activity data is not accompanied with their sleep ground truth. It uses the time series of phone activity features, $X^k = \{x_1, x_2....x_T\}$, where x_T represents each time step, into $Model_{Teacher}$ to predict whether user k is asleep or awake. The output generated by $Model_{Teacher}$ constitute *pseudo labels* for the user.

It is possible that $Model_{Teacher}$ does not generalize well to the new user, likely due to low-quality pseudo labels. We use Dropout in the prediction network to improve the reliability of pseudo labels by reducing the effects of overfitting and improving the robustness of the model's predictions. Dropout works by randomly dropping out some neurons in the network during each training iteration. We also calibrate our models using a validation set. *SleepLess* performs label selection to only retain output predictions of high confidence, discarding all others. We use a confidence threshold, Δ, retaining predictions above this value as pseudo labels for the next phase.The Confidence score is chosen based on the average softmax scores of all predicted outcomes in a given 24-h period as follows:

$$C^i_{avg} = \frac{\sum c^i_t}{T} \tag{2}$$

Thus, for each new user, we obtain pseudo labels $\hat{Y} = \{\hat{y}_1, \hat{y}_2 .... \hat{y}_T\}$. Not all time slots will have such labels but since it is easy to collect unlabeled data, this process can continue until adequate pseudo labels are generated.

Step 3: Personalize a Student Model Using Fine-Tuning. The pseudo label data for user k can then be used to train a personalized model. Specifically, *SleepLess* performs transfer learning of the original teacher model, $Model_{Teacher}$, by freezing the initial layers of the CNN. The pseudo labels are then used to further train and fine-tune the subsequent layers, thus generating a student model, $Model_{Student}$, now personalized to the new user. This transfer learning approach helps the model re-use the learnt features in earlier layers and tailor the latter layers to specific sleep patterns of the new user.

5 *SleepLess* Implementation

We implemented *SleepLess* prototype as a cloud-based service using Python [27] and Keras libraries [28]. In what follows, we elaborate on the steps we took to build the deep learning component of the system.

5.1 Data Pre-processing

Network Activity Data. In essence, our technique utilizes users' phone activities generated from smartphone devices to predict their sleep. These activities can be represented by various measures including but not limited to the screen activity, application logs, accelerometer, and WiFi network activity logs [13].

Our work builds on utilizing WiFi network activity data for several operational reasons. As discussed in Sect. 6.1, our study sought to minimize user burdens and maximize privacy by avoiding dedicated app installations on their smartphone and, thus, directly sensing from their device. As such, we use a passive sensing technique where we acquired WiFi network activity data to collect *syslog* data directly from the WiFi APs, bypassing any connection to the user's device. In acquiring these logs, we filter out entries relevant only to our participants, specifically their primary smartphone device. The coarse granularity of WiFi data presents inherent technical errors in the measurement instrument. The coarse granularity of WiFi data presents inherent technical errors in the measurement instrument. To maintain the quality of our analysis, we cleaned these logs to reduce inaccurate and noisy data.

5.2 Feature Extraction

SleepLess processes logs of WiFi network activity rates generated from a user's phone. These logs are in the following format:

```
<date> <hh:mm:ss> <controller> <event\_ID> <severity>

<AP, MAC and IP addresses> <message text with BSSID and SSID>
```

Timestamp is given by date and time, while WiFi access point(AP) and users' device MAC addresses. Note that our collection of WiFi logs ensures user privacy by hashing users' device MAC addresses. together allow us to approximate the user's location. Event ID particularly describes three events of interest. They are i) association, when a device connects to an access point , ii) dissociation, when a device disconnects from the access point, and iii) authentication of the device, thus allowing us to approximate user activity and movement from one place to another. The result of using a *timestamp*, *event_ID*, *WiFi AP address*, and user device (hashed) *MAC address* is four input features to predict the user's sleep.

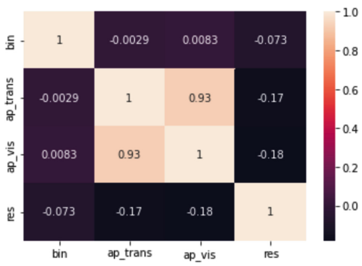

Fig. 3. Feature correlation

Table 1. Features used

Weight	Feature
0.647	Time
0.107	#WiFi AP Connections
0.132	#WiFi AP transitions
0.098	Dorm or not

- The *time of day* is marked in bins of every 15 min. Since we are interested in the nocturnal sleeping period, we consider a 24-h time span that starts from 6 pm of Day_1 and ends at 5:45 pm the next day. 6 pm of Day_1 corresponds to bin 0, while 5:45 pm of Day_2 corresponds to bin 95.
- The *number of WiFi AP connections* denotes the total number of unique access points visited over every 15 min interval.

– The *number of WiFi AP transitions* denotes the total number of transitions approximated from WiFi AP switching over every 15 min interval.
– We categorize *Dorm or not* as a user in their residential or non-residential location. This assignment is based on mapping WiFi APs specific to our campus and campus housing.

Figure 3 shows the correlation map of our features. Then, we compute feature importance through the permutation importance method. That is, we recursively measure the model performance every time the values of a feature are randomly shuffled. Table 1 summarizes the values of the most important features in our model. In this case, our top two features represent the time at which the user's device generate high network activity rate.

5.3 Model Architecture

In developing the teacher model, $Model_{Teacher}$, we extract features from the time series data of a fixed set of users into bins of 15 min intervals and include label assignments of *sleep (1)* or *awake (0)* state corresponding to their supplied ground truth. Conversely, $Model_{Student}$ is accompanied by pseudo labels of sleep.

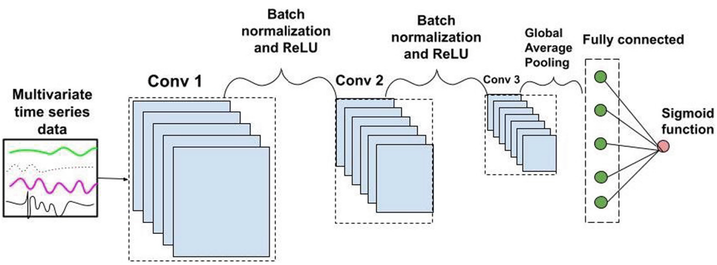

Fig. 4. CNN architecture for our Teacher-Student models.

We utilize a basic CNN architecture, depicted in Fig. 4. Specifically, the model consists of three temporal convolution layers with a filter size of 32, 64, and 96. Each layer has a kernel size of 24, 16, and 8, respectively. We chose a uniform stride of 1 for all the layers. The ReLU function was chosen as the activation function. We also chose a dropout rate of 0.1 between layers, a global 1D maximum pooling layer at the last convolutional layer. Finally, we used Adam optimizer with a learning rate of 0.0003.

6 Evaluation of *SleepLess*

In this section, we evaluate the efficacy of *SleepLess*'s semi-supervised learning model with other prediction methods employed in similar prior work. Further, we evaluate its robustness on private home users with different phone activity profiles.

Table 2. Demographic information of two different datasets.

	Student dataset (main)	Private home dataset (supplementary)
Users, N	20 (18 Male, 2 Female)	3 (2 Male, 1 Female)
Age	18–21 years old mean: 20, stdev.: 0.75	36–46 years old mean: 42, stdev.: 4.71
Study duration	4 weeks	1 week–4 weeks
Sleep summary	Bedtime: 06:00 pm–11:00 am (mean: 01:20 am), Wake time: 03:00 am–03:00 pm (mean: 10:10 am), Sleep duration: 60–660 min (mean: 420 min)	Bedtime: 11:20 pm–12:45 am (mean: 11:27 pm), Wake time: 05:30 am–08:00 am (mean: 06:16 am), Sleep duration: 300–511 min (mean: 428 min)
Sleep Tracker	Fitbit Inspire HR	Fitbit Inspire HR, Fitbit Versa 3 and manual logs
Device activity	anonymized logs of connected smartphones to campus WiFi.	unanonymized WiFi logs of connected smartphones and home devices to home WiFi

6.1 Experimental Setup

We begin by describing our user study details in acquiring two different datasets, as summarized in Table 2.

Study Protocol. We ran a month-long user study among college students living on campus upon receiving IRB approval from our institution. Our study protocol includes recruiting undergraduates and giving out Fitbit inspire HR [10] wearable to collect sleep logs automatically. The sleep logs generated from Fitbit are used as ground truth. In practice, WiFi logs generated from the campus APs only contain the timestamp and network activity rate of hashed MAC addresses per device. We specifically isolated our participants' device connection to a dedicated AP to identify our participant's smartphones, despite us dealing with only hashed records. Separately, we repeated the same protocol to a different set of private home users (non-student) over a one-week period.

Data Ethics and IRB Approval. Our user study is approved by the Institutional Review Board (IRB) and includes a Data Usage Agreement (DUA) with the campus network IT group.

6.2 Efficacy of SSL-Based Model

Our first experiment examines how *SleepLess* performs compared to a personalized supervised learning approach, which would require re-training a small amount of labeled data for a completely new user.

We utilize the student dataset in this experiment. Our results are achieved through conducting a train-test split. Specifically, models are trained on ten randomly picked student users and tested on the remaining ten users. Additionally, we set aside the first two weeks of labeled data of our test users to develop their personalized model and used the last two weeks to test the model.

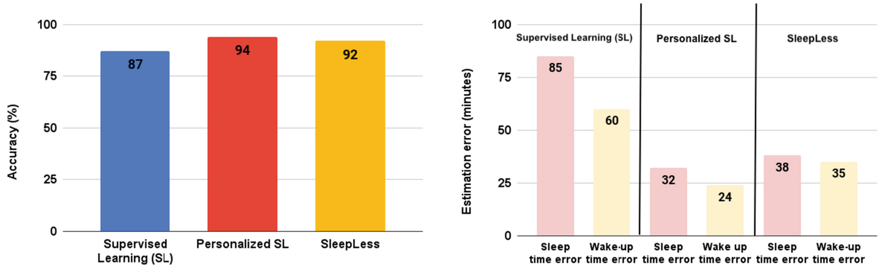

Fig. 5. Model performance comparing general and personalized supervised learning approaches with semi-supervised learning that *SleepLess* employs.

As shown in Fig. 5, the model performance for *SleepLess* is comparable to that of a personalized model. Specifically, the sleep and wake times prediction errors by *SleepLess* are approximately 10 min more than a personalized model. As expected, personalization will yield better model performance for a new user (94% accuracy, 32 min sleep time prediction error and 24 min wake-up time prediction error). In contrast, *SleepLess* achieves 92% accuracy, 38 min sleep time prediction error and 35 min wake-up time prediction error.

Table 3. Personalized supervised learning versus semi-supervised learning on home users.

Method	T_{sleep} min	T_{wake} min
Personalized SL	8 ± 5	23 ± 20
SleepLess	12 ± 4	25 ± 20

We replicated this experiment on our small group of non-student users who reside in private homes, comparing a personalized supervised learning approach with *SleepLess*. As indicated in Table 2 of our user study, the device network activity of our home participants accounted for all the personal devices connected to their home WiFi AP. Our model yielded slightly increased errors in predicted sleep and wake times. Even though *SleepLess* recorded 2% less accuracy, our results remain favorable for two reasons. First, *SleepLess*'s model decrement is insignificant ($p < .01$). Second, personalized supervised learning model will only offer practical use to new users after providing two weeks of labeled data for model retraining. In a real-world application, *SleepLess* can begin prediction for a new user without their labeled data almost instantaneously. These considerations motivate us to explore and propose *SleepLess* as a more practical approach, appealing to new users.

6.3 *SleepLess* vs. Baseline Algorithms

We follow up on the comparison of *SleepLess*'s performance with other baseline approaches proposed in prior work, including why these baselines would work less favorably for our case.

Baseline Algorithms. Given that the primary motivation of our work is to utilize fewer labeled data, we selectively picked on several methods that sought to learn with fewer labels.

The first is *semi-supervised learning using self-training*. Here, we train the teacher model using the labeled data from existing users and generate pseudo labels for the new users using their unlabeled data. The pseudo labels are picked based on the same criterion we used in *SleepLess*. The selected pseudo labels are combined with labeled data, D to train the student model. Compared to the typical self-training process, we stopped the pseudo label selection after one iteration to avoid error accumulation.

The second is *multi-head single-view co-training*. We adopt a similar structure suggested by Chen et al. [14], with several tweaks to the training pipeline. Rather than training multiple classifiers as in multi-view co-training, we will use predictions from multiple classification heads sharing a common module. First, we train the classifier using the labeled data, D, with only one classification head. Second, we generate pseudo labels for the new user using major voting by all the classification heads. Similarly, we filter the pseudo labels using the label selection criteria as per *SleepLess*. Finally, we combine the selected pseudo labels and labeled data to train the personalized model for the new user.

The third baseline employs an *encoder-decoder* approach, which exploits unlabeled data to learn the latent representation of the data [29]. The central idea of this approach is that the unlabeled data and labeled data can together help us select relevant features thus improving model robustness and generalizability. Here, we first combine unlabeled data from the new user to labeled data from existing users and train an encoder-decoder model to learn the latent representation of the new users' unlabeled data. After training the encoder-decoder model, we ingest only the labeled data, D, through the encoder decoder and obtain intermediate output from the encoder. We train the classifier using the encoded representation.

Results. Table 4 compares *SleepLess*'s model performance against the baseline models by testing on 4 weeks of data from 10 users. Specifically, *SleepLess* yields 96% accuracy, which is significantly higher than all other models ($p < .01$). Further, Fig. 6 charts the sleep and wake time errors for all methods. We observed that multi-head single-view co-training yielded the least errors among all the baseline methods; 27 min of sleep time error and 18 min of wake time error. However, this difference remains significantly higher than *SleepLess*. Co-training, which relies on the mix of new and existing user data, can be more suitable in conditions where our goal is to improve a generalized model approach.

Table 4. *SleepLess* and baseline models performance.

Method	Acc	Prec	Rec	F Scr.	p val.
SleepLess- SSL	0.96	0.98	0.87	0.93	–
Self-training	0.91	0.94	0.88	0.86	$p < .01$
Multi-head single-view Co-Training	0.92	0.87	0.88	0.88	$p < .01$
Auto encoder	0.87	0.80	0.76	0.78	$p < .05$

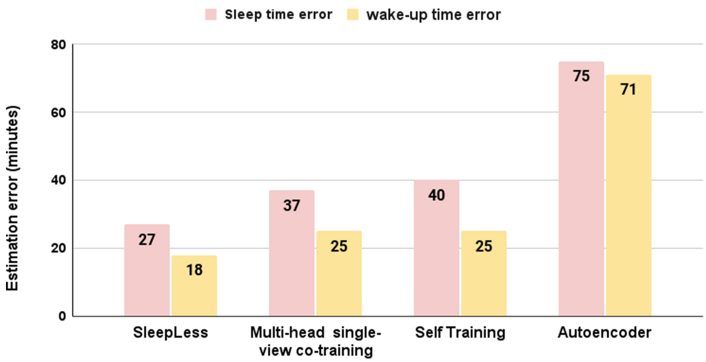

Fig. 6. Sleep and wake time errors by *SleepLess* and baselines.

Key Takeaway. Given our proposal on a semi-supervised learning approach, we compared the performance of our technique with standard SSL-based techniques such as self-training and co-training. As briefly discussed, our decision to compare these standard techniques was informed by prior work's implementation for sleep prediction, however, using fine-grained data. In this case, the consideration for a teacher-student model is primarily to address the error accumulation problem, which we hypothesized will be more prominent from using coarse-grained data such as phone network activity rate. Our results yielded significantly better performance of 96% accuracy compared to these standard techniques.

7 Conclusions

Fundamentally, the requirement of collecting a significant amount of ground truth holds for training any user behavioral models. Unlike many prior sleep detection techniques that rely on collecting and training large amounts of labeled data, *SleepLess*, w uses semi-supervised learning over unlabeled data sensed from the user's smartphone network activity to develop personalized models and detect their sleep duration for the night. By using a generalized pre-trained model on an existing set of users to produce pseudo labels for unlabeled data of a new user, it achieves personalization by fine-tuning using selected pseudo-labels for the new user without requiring any labeled data. Our user study among 23 users found *SleepLess* model yielding around 96% accuracy, between 12–27 min of sleep time error and 18–25 min of wake time error. With our prediction technique yielding the best performance, our work shows promise for sleep monitoring to be more conveniently adapted to monitor new users' sleep immediately. Where the larger goal of our work aims to improve students' health, lack of sleep is linked to many major health challenges. Our work continues investigating the efficacy of this technique in complementary domains, including sleep quality.

Acknowledgments. This work is supported in parts by NSF grants 2211302, 2211888, 2213636, 2105494, and Army grant W911NF-17-2-0196.

References

1. Krueger, P.M., Friedman, E.M.: Sleep duration in the United States: a cross-sectional population-based study. Am. J. Epidemiol. **169**(9), 1052–1063 (2009)
2. Rahman, T., et al.: Dopplesleep: a contactless unobtrusive sleep sensing system using short-range doppler radar. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (2015)
3. Abdullah, S., 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)
4. De Zambotti, M., et al.: Wearable sleep technology in clinical and research settings. Med. Sci. Sports Exercise **51**(7), 1538 (2019)
5. Ohayon, M.M.: Epidemiological overview of sleep disorders in the general population. Sleep Med. Res. **2**(1), 1–9 (2011)
6. Nest. https://support.google.com/googlenest. Accessed 27 Oct 2022
7. Amazon. https://www.amazon.com/Paschar-LLC-Walabot-Sleep-Tracker/dp/B07C2HRYSX. Accessed 27 Oct 2022
8. Zhao, Y., et al.: Semi-supervised federated learning for activity recognition. arXiv preprint arXiv:2011.00851 (2020)
9. Trivedi, A., et al.: Wifitrace: network-based contact tracing for infectious diseases using passive wifi sensing. In: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies vol. 5, no. 1, pp. 1–26 (2021)
10. Google. https://www.fitbit.com/global/be/products/-trackers/inspire. Accessed 27 Oct 2022
11. Wang, R., et al.: StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (2014)
12. Hofman, J.M., Sharma, A., Watts, D.J.: Prediction and explanation in social systems. Science **355**(6324), 486–488 (2017)
13. Saeb, S., et al.: Scalable passive sleep monitoring using mobile phones: opportunities and obstacles. J. Med. Internet Res. **19**(4), e118 (2017)
14. Chen, M., et al.: Semi-supervised learning with multi-head co-training. Proc. AAAI Conf. Artif. Intell. **36**(6), 6278–6286 (2022)
15. Zhang, C., et al.: CMS2-net: semi-supervised sleep staging for diverse obstructive sleep apnea severity. IEEE J. Biomed. Health Inf. **26**(7), 3447–3457 (2022)
16. Li, Y., et al.: Adversarial learning for semi-supervised pediatric sleep staging with single-EEG channel. Methods **204**, 84–91 (2022)

17. Haoran, B., Guanze, L.: Semi-supervised end-to-end automatic sleep stage classification based on pseudo-label. In: 2021 IEEE International Conference on Power Electronics, Computer Applications (ICPECA). IEEE (2021)
18. El-Khadiri, Y., et al.: Sleep activity recognition using binary motion sensors. In: 2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI). IEEE (2018)
19. Cuttone, A., et al.: Sensiblesleep: a Bayesian model for learning sleep patterns from smartphone events. PloS One **12**(1), e0169901 (2017)
20. Peng, Y., et al.: Joint semi-supervised feature auto-weighting and classification model for EEG-based cross-subject sleep quality evaluation. In: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE (2020)
21. Heremans, E.R.M., et al.: From unsupervised to semi-supervised adversarial domain adaptation in electroencephalography-based sleep staging. J. Neural Eng. **19**(3), 036044 (2022)
22. Min, J.-K., 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)
23. Hao, T., Xing, G., Zhou, G.: isleep: unobtrusive sleep quality monitoring using smartphones. In: Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems (2013)
24. Ren, Y., et al.: Fine-grained sleep monitoring: Hearing your breathing with smartphones. In: 2015 IEEE Conference on Computer Communications (INFOCOM). IEEE (2015)
25. Chen, Z., et al.: Unobtrusive sleep monitoring using smartphones. In: 2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops. IEEE (2013)
26. Gu, W., et al.: Sleep hunter: towards fine grained sleep stage tracking with smartphones. IEEE Trans. Mobile Comput. **15**(6), 1514–1527 (2015)
27. Python. https://www.python.org/. Accessed 27 Oct 2022
28. keras. https://keras.io/. Accessed 27 Oct 2022
29. Bhattacharya, S., et al.: Using unlabeled data in a sparse-coding framework for human activity recognition. Pervas. Mobile Comput. **15**, 242–262 (2014)
30. Munk, A.M., et al.: Semi-supervised sleep-stage scoring based on single channel EEG. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE (2018)
31. Wang, X., et al.: Smartphone sonar-based contact-free respiration rate monitoring. ACM Trans. Comput. Healthc. **2**(2), 1–26 (2021)
32. Tiron, R., et al.: Screening for obstructive sleep apnea with novel hybrid acoustic smartphone app technology. J. Thoracic Dis. **12**(8), 4476 (2020)
33. Kim, D.H., Kim, S.W., Hwang, S.H.: Diagnostic value of smartphone in obstructive sleep apnea syndrome: a systematic review and meta-analysis. PloS One **17**(5), e0268585 (2022)
34. Goldblum, M., et al.: Dataset security for machine learning: data poisoning, backdoor attacks, and defenses. IEEE Trans. Pattern Anal. Mach. Intell. **45**(2), 1563–1580 (2022)
35. Dixon, M., et al.: Sleep-wake detection with a contactless, bedside radar sleep sensing system (2021)
36. Rundo, J.V., Downey, R., III: Polysomnography. Handb. Clin. Neurol. **160**, 381–392 (2019)

37. Blunck, H., et al.: On heterogeneity in mobile sensing applications aiming at representative data collection. In: Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication (2013)
38. Wuzheng, X., et al.: Semi-supervised sparse representation classification for sleep EEG recognition with imbalanced sample sets. J. Mech. Med. Biol. **21**(05), 2140006 (2021)
39. ouraring. https://ouraring.com/. Accessed 27 Oct 2022
40. Apple. https://www.apple.com/watch/. Accessed 27 Oct 2022
41. Witt, D.R., et al.: Windows into human health through wearables data analytics. Curr. Opin. Biomed. Eng. **9**, 28–46 (2019)
42. Lee, Y.S., et al.: Monitoring and analysis of respiratory patterns using microwave doppler radar. IEEE J. Trans. Eng. Health Med. **2**, 1–12 (2014)
43. Gu, C., Li, C.: Assessment of human respiration patterns via noncontact sensing using doppler multi-radar system. Sensors **15**(3), 6383–6398 (2015)
44. Lin, F., et al.: SleepSense: a noncontact and cost-effective sleep monitoring system. IEEE Trans. Biomed. Circuits Syst. **11**(1), 189–202 (2016)
45. Hong, H., et al.: Microwave sensing and sleep: Noncontact sleep-monitoring technology with microwave biomedical radar. IEEE Microw. Magaz. **20**(8), 18–29 (2019)
46. Zakaria, C., et al.: SleepMore: inferring sleep duration at scale via multi-device WiFi sensing. In: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 6, no. 4, pp. 1–32 (2023)
47. Zhu, X., Goldberg, A.B.: Introduction to Semi-supervised Learning. Springer Nature (2022)
48. Zhu, X.J.: Semi-supervised learning literature survey (2005)
49. Miyato, T., et al.: Virtual adversarial training: a regularization method for supervised and semi-supervised learning. IEEE Trans. Pattern Anal. Mach. Intell. **41**(8), 1979–1993 (2018)
50. Yalniz, I.Z., et al.: Billion-scale semi-supervised learning for image classification. arXiv preprint arXiv:1905.00546 (2019)
51. Zou, Y., et al.: Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In: Proceedings of the European Conference on Computer Vision (ECCV) (2018)
52. Zhang, C., et al.: Understanding deep learning (still) requires rethinking generalization. Commun. ACM **64**(3), 107–115 (2021)
53. Zhou, Z.-H., Li, M.: Semi-supervised learning by disagreement. Knowl. Inf. Syst. **24**, 415–439 (2010)
54. Wang, W., Zhou, Z.-H.: Analyzing co-training style algorithms. In: European Conference on Machine Learning. Springer, Heidelberg (2007)
55. Guan, D., et al.: Activity recognition based on semi-supervised learning. In: 13th IEEE International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA 2007). IEEE (2007)
56. Faridee, A.Z.M., et al.: Strangan: adversarially-learnt spatial transformer for scalable human activity recognition. Smart Health **23**, 100226 (2022)
57. Gogna, A., Majumdar, A.: Semi supervised autoencoder. In: Neural Information Processing: 23rd International Conference (ICONIP 2016), Kyoto, 16–21 October 2016, Proceedings, Part II, vol. 23. Springer (2016)
58. Chakma, A., et al.: Activity recognition in wearables using adversarial multi-source domain adaptation. Smart Health **19**, 100174 (2021)
59. Pearce, T., Brintrup, A., Zhu, J.: Understanding softmax confidence and uncertainty. arXiv preprint arXiv:2106.04972 (2021)

60. Dhamija, A.R., Günther, M., Boult, T.: Reducing network agnostophobia. Adv. Neural Inf. Process. Syst. **31** (2018)
61. Japkowicz, N., Stephen, S.: The class imbalance problem: a systematic study1. Intell. Data Anal. **6**(5), 429–449 (2002). https://doi.org/10.3233/IDA-2002-6504