IPSJ SIG Technical Report

Vol.2017-MBL-82 No.54
Vol.2017-UBI-53 No.54
2017/3/10

Inter-Individual Differences in Sleep Quality: Insights from Mining Wearable Sleep-Tracking Data

ZILU LIANG[†1] MARIO ALBERTO CHAPA-MARTELL[†2] BERND PLODERER[†3]

Abstract: Many people are using mobile and wearable technologies to collect their sleep data in order to learn more about their sleep patterns and to identify sleep anomalies. However, these sleep data are rarely analyzed in a systematic way and many users find it difficult to understand if their sleep patterns are normal. To address this issue, we applied a data mining technique to analyze personal sleep tracking data collected from 7 participants, aiming to investigate inter-individual difference in terms of abnormal sleep patterns. We found that population criterion may classify people who are satisfied with their sleep experience as poor sleepers. On the other hand, sleep anomalies are missed for people whose personal sleep distribution deviates from sleep distribution of the population. These abnormal sleep patterns manifested themselves in different dimensions. For example, one person may have problems in sleep duration, while another person may sleep enough but have lots of awakenings. Even within one dimension, abnormal sleep patterns of one person can be antithetical to those of another person. The quantitative analysis results on personal sleep data support the conclusions of previous qualitative studies on sleep tracking, suggesting the need for a change in evaluating sleep quality that takes into account interpersonal differences.

Keywords: Wearable computing; Sleep; Health; Personal informatics; Multivariate outlier detection

1. Introduction

Good sleep is essential for personal health, while poor sleep can negatively affect our wellbeing and lead to illness [1]. In the past, sleep monitoring could only be done in sleep clinics using Polysomnography (PSG) which is expensive and obtrusive. Recently, sleep-tracking has become more accessible to people in their own homes to the advances in affordable and easy-to-use mobile and wearable technologies life the Fitbit wristband [2]. A screenshot of the Fitbit dashboard (see Figure 1) shows that users can learn how long they sleep and interruptions to their sleep. According to previous validation studies, Fitbit generates reasonably reliable measurements and has been widely used in fields to assess sleep measures [3-4].

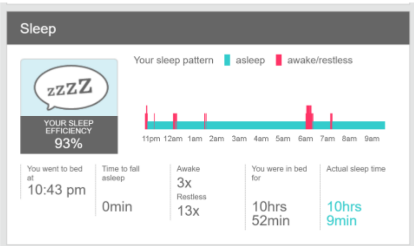

Figure 1 A screenshot of the sleep-tracking data presented on Fitbit dashboard.

One envisioned application of these home sleep monitoring technologies is to detect abnormal sleep patterns. For patients, abnormal sleep patterns may suggest side-effects of medical interventions; while for healthy people, detecting abnormal sleep patterns can alert them of possible sleep disorders and related diseases. However, such kind of data analysis is not available in current commercial sleep-trackers, because information is missing on what constitutes a normal sleep pattern. Our previous study showed that users find it difficult to understand the data they collect, let alone if their sleep habits are healthy [5]. People asked, for example, if 10 hours sleep are normal; what it means if they have 13 periods of restlessness, and how they can integrate sleep with other data to get a holistic understanding of their lifestyle? Indeed, it is difficult to understand sleep because we do not consciously experience it, and because it is difficult to recollect past nights and detect patterns over days and weeks. This gap between sleep data and personal insights calls for the design and development of advanced and automated data analysis approaches and tools.

In this study, we applied a data mining technique to analyze personal sleep-tracking data with the purpose of investigating inter-individual difference in terms of abnormal sleep. Defining "abnormal" as the statistical outliers from each user's own sleep baseline, we formulated the problem as an outlier detection problem. Since sleep is multidimensional in nature and is characterized by several metrics [7], a multivariate outlier detection technique was employed to achieve the goal. Fitbit can track six sleep metrics, i.e., minutes asleep (MASP), minutes awake (MAWK), number of awakenings (NAWK), bed time (BT), minutes to fall asleep (MTFA) and sleep efficiency (SE). The last metric was discarded since it is a redundant feature of MASP and MAWK.

We recruited 7 participants and collected sleep data using Fitbit wristbands. A multivariate outlier detection data mining technique was applied to the collected sleep-tracking data to detect abnormal sleep patterns for each participant. We found that population criterion may classify people who were satisfied with their sleep experience as poor sleepers, while fail to detect sleep anomalies for people whose personal sleep distribution deviated from sleep distribution of the population. In comparison, the data mining technique successfully detected abnormal sleep for each individual by counting in their personal sleep quality baselines. The detected abnormal sleep patterns demonstrated rich diversities. On the one hand, abnormal sleep

†1 The University of Tokyo.
†2 CAC Corporation. The opinions expressed in this paper are solely of the author's and do not represent the views of the company.
†3 Queensland University of Technology.

manifested in multiple dimensions, and the dimensions were not necessarily equivalent to all people. For another, even in a same dimension, abnormal sleep patterns of one person could be antithetical to those of another. In summary, the “abnormal” could mean either short or long sleep duration, either early or late time to go to bed, more awakenings during sleep, or long sleep onset latency. The outcome of this study echoes previous studies that normal sleep is hard to define because people differ in physiology, psychology, lifestyle and living environment [5,7], suggesting the need for a paradigm shift in sleep quality evaluation approach that counts in interpersonal differences.

2. Related Work

2.1 Quantified self and sleep-tracking technologies

Under the influence of the Quantified Self movement, many people are using wearable or portable sensors to track their physiological (e.g. weight, blood pressure, blood glucose level) and psychological (e.g. mood, stress, emotion) parameters for better health and wellbeing [17-19]. Among various health aspects, sleep has attracted wide attention from quantified-self followers, largely due to the importance of sleep to our overall health [20-21] and also to the prevalence of sleep problems in modern society [22-23]. In recent years, a large number of commercial sleep-tracking technologies appeared in consumer market. Depending on the underline technologies, sleep trackers could be classified into three groups: mobile applications (e.g. SleepAsAndroid, SleepBot), wearable devices (e.g. Fitbit, Neuroon) and embedded systems (e.g. Beddit). They provide users with information such as total sleep duration (e.g. minutes asleep), sleep continuity (e.g. minutes awake, awakening counts), sleep onset latency (e.g. minutes to fall asleep), and even sleep stages (e.g. REM sleep and slow-wave sleep). Devices such as Jawbone also provide tailored coaching tips on how to promote healthy sleep habits for better sleep quality.

Although these technologies are promising for home sleep monitoring in terms of convenience and cost, several challenges remains [8-9]. Two of the most mentioned problems are data quality issue and data interpretation. Whereas there is increasing evidence that commercial wearable devices produce reasonably accurate measures on sleep [3-4], it remains a challenge as to how to empower users to gain insights from their sleep-tracking data.

2.2 Sleep structure and sleep health

Human sleep is multidimensional and can be characterized from the perspective of quantity, continuity and timing [10]. In sleep research, sleep quality is evaluated using a set of metrics called sleep structure [7] and clinical sleep test usually involves monitoring the whole sleep structure. According to [7], the most relevant dimensions of sleep health include the following five aspects.

- Sleep duration: the total amount of sleep obtained per 24 hours.
- Sleep continuity or efficiency: the ease of falling asleep and returning to sleep.
- Timing: the placement of sleep within the 24-hour day.
- Alertness/sleepiness: the ability to maintain attentive wakefulness.
- Satisfaction/quality: the subjective assessment of “good” or “poor” sleep.

Many modern sleep-tracking technologies can track several sleep metrics that fit in the above five aspects. The values of these sleep metrics are presented to users on a dashboard on daily basis. Since detecting high-dimensional anomalies through human observation and reasoning is challenging, if not impossible, users of self-tracking technologies found it difficult to gain insights from their sleep data. In most cases, users would simply focus on one or two metrics that are most relevant to their sleep quality [5]. We argue that the value of multi-dimensional sleep data should not be compromised due to the complexity of analysis. Advanced data analysis techniques may help establish larger personal value from self-tracking sleep data. That is the motivation for us to conduct this study.

3. Methods

3.1 Collecting sleep-tracking data

We used the sleep-tracking data collected in our previous study [5] from 7 participants (5 female, year range = 20~50 years old). These participants were recruited through university mailing list and they had been using Fitbit to track their sleep. We obtained informed consent from all participants. Their sleep quality during the past month was evaluated using Pittsburg Sleep Quality Index (PSQI) [11]. A PSQI over 5 indicates poor sleep quality. Participant #3 was diagnosed with sleep apnea and was under treatment throughout the study. Participant #4 had visited a sleep clinic before but did not have severe sleep problems during the study. The demographic information of the participants was summarized in Table 1.

Table 1. Demographic information of participants.

ID	Age Range	Type of Fitbit	PSQI Baseline
#1	20s	Charge HR	1 (Good)
#2	30s	Charge HR	3 (Good)
#3	40s	Charge HR	4 (Good)
#4	40s	Flex	11 (Poor)
#5	30s	Charge	7 (Poor)
#6	30s	Charge HR	5 (Poor)
#7	30s	Flex	4 (Good)

Participants used their own Fitbit devices to collect sleep data during the study, which included Fitbit Flex, Fitbit Charge, and Fitbit Charge HR. The first two models require users to manually switch in and out of sleep-tracking mode, while Fitbit Charge HR automatically switch tracking mode. In total 7 datasets were collected, and each dataset has 34 ~ 61 data entries depending on the participant. The prepared datasets satisfied the 30-day requirement for personal data analysis [12].

3.2 Mining sleep-tracking data

3.2.1 Problem formulation

In this study, the goal of the data mining task was to detect abnormal sleep from personal sleep-tracking data collected using Fitbit. Since these data were collected in free conditions and were not labeled, we used an unsupervised outlier detection technique to perform the data mining task. We adopted a general approach to define “abnormal” as the statistical outliers from each user’s own sleep baseline. The detection of abnormal sleep was thus formulated into the following multivariate outlier detection problem.

Given D days of n-dimensional sleep-tracking data $\boldsymbol{X} = \{\boldsymbol{x_d}\}_{d=1}^{D}$ where $\boldsymbol{x_d} = (x_{d,1}, x_{d,2}, \dots, x_{d,n})$, find $\boldsymbol{X_{unusual}} = \{\boldsymbol{x_u}\}_{u=1}^{U} \subset \boldsymbol{X}$ that satisfies (1).

$$|\boldsymbol{x_u} - \boldsymbol{x_0}|/\xi > t \quad (1)$$

where $\boldsymbol{x_0}$ is the center of the dataset, $|\boldsymbol{x_u} - \boldsymbol{x_0}|$ represents the distance between $\boldsymbol{x_0}$ and $\boldsymbol{x}$, ξ is the natural variation of the dataset and t is a threshold parameter.

Fitbit tracked six sleep metrics, i.e., minutes asleep (MASP), minutes awake (MAWK), number of awakenings (NAWK), bed time (BT), minutes to fall asleep (MTFA) and sleep efficiency (SE). The last metric was discarded since it is a redundant feature of MASP and MAWK. Therefore, the sleep-tracking data used for data mining had 5 dimensions.

3.2.2 Detecting abnormal sleep from multivariate data

As is shown in Equation (1), we need to quantify the distance between a multivariate data point and the reference center. We used the standard method, i.e. Mahalanobis distance [13], to perform the task. Assuming that the data satisfied multivariate normal distribution with estimated mean μ and covariance matrix Σ, the Mahalanobis distance $d_M(x)$ measures how many standard deviation away a data point $\boldsymbol{x}$ is from the mean.

$$\mathrm{d}_M(\boldsymbol{x})^2 = (\boldsymbol{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \quad (2)$$

where $\boldsymbol{x} = (x_{d,1}, x_{d,2}, x_{d,3}, x_{d,4}, x_{d,5})^T$ is a set of 5-dimension sleep data points, $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3, \mu_4, \mu_5)^T$ is the mean of the distribution and Σ is the covariance matrix. Since $\boldsymbol{\mu}$ and Σ are very sensitive to outlier as well, we used Minimum Covariance Determinant estimator [14] to robustly estimate them. The 1% quantile of $\mathrm{d}_M(\boldsymbol{x_d})$ was taken as outliers

4. Results

The distinct characteristics of identified abnormal sleep for each participant were summarized in Table 2. The detected abnormal sleep manifested in multiple dimensions, and the dimensions were not necessarily equivalent to all participants. For example, sleep duration was irrelevant to participants #4 and #6. In a similar way, awakening duration was irrelevant to participants #5 and #6, awakening count was irrelevant to participants #2, #3 and #5, and sleep onset latency was irrelevant to participants #1, #2 and #5. Bed time was the only metric that characterizes abnormal sleep for all participants in the study cohert.

The average sleep duration of usual and abnormal sleep for each participant was plotted in Figure 2. The dash lines represent the recommended sleep duration by sleep research community in the dimension of sleep duration [26]. The 6-9-hour standard would classify participants #2, #5, #6, #7 as good sleepers and participant #3 as poor sleeper, and may not able to detect personal abnormal sleep patterns for these people as all their sleep appear to be homogenously good or poor to the standard. In practice, however, participant #3 considered himself as a good sleeper, while participants #5 and #6 were not satisfied with their sleep quality.

Table 2. Characteristics of abnormal sleep for each participant

ID	*Characteristics of Abnormal Sleep*				
	Sleep Duration	Awakening Duration	Awakening Count	Bed Time	Sleep Onset Latency
#1	Short	Long	High	Late	/
#2	Short	Long	/	Late	/
#3	Short	Long	/	Late	Long
#4	/	Long	High	Early	Long
#5	Short	/	/	Late	/
#6	/	/	High	Early	Long
#7	Long	Long	High	Late	Long

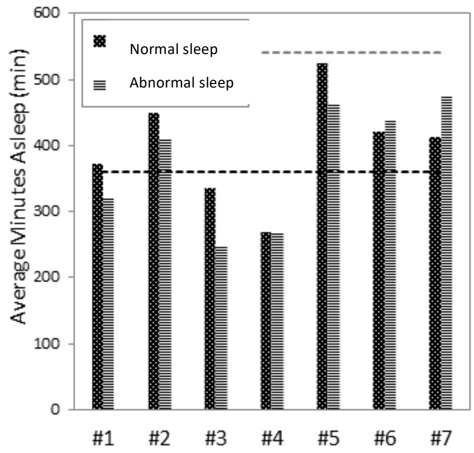

Figure 2 Average sleep duration of usual and abnormal sleep for each participants.

The average time to go to bed was plotted in Figure 3. The time format was converted to numerical data for analysis; 2050 represents 20:50. The difference of bed time between usual and abnormal sleep was more than 1 hour for good sleepers (i.e. participants #1, #2, #3, #7). Out of the 7 participants, abnormal sleep was characterized by early bed time for 2 participants (#4 and #6) while by late bed time for the rest. Coincidently participants #4 and #6 were both poor sleepers according to their PSQI baseline. We also noticed that the abnormal sleep patterns of the 2 poor sleepers share the common characteristics of early bed time (Figure 3) and long sleep onset latency (Figure 5), hypothesizing a link between the two phenomenon among poor sleepers.

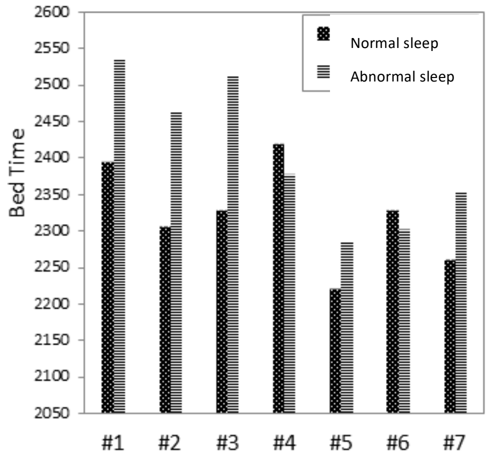

Figure 3 Average bed time of usual and abnormal sleep for each participants. (Time format was converted to numerical data for analysis; 2050 represents 20:50)

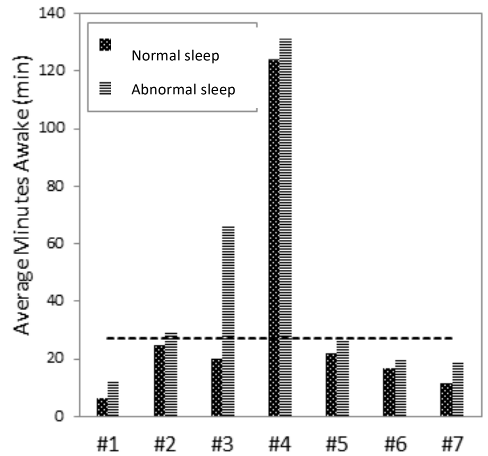

Figure 4 Average awake duration of usual and abnormal sleep for each participants.

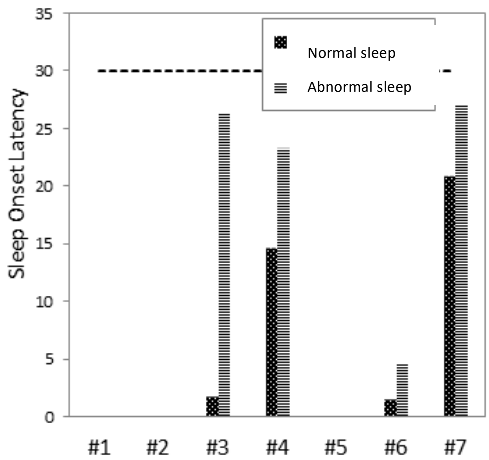

Figure 5 Average sleep onset latency of usual and abnormal sleep for each participants.

The average awake duration of usual sleep and that of abnormal sleep was plotted in Figure 4. The dash line represented the population criterion on awakening duration, i.e. being awake for longer than 30 minutes was considered abnormal. The population criterion was only valid for participants #2 and #3, where the awakening duration of usual sleep was below 30 minutes and vice versa. It would not be possible to detect personal abnormal sleep for the rest of the participants using the population criterion, as the awakening durations of usual and abnormal sleep were all above 30 minutes for participants #4 and all below 30 minutes for participants #1, #5, #6, #7. Similarly, the population criterion of 30 minutes was not able to detect anomaly in terms of sleep onset latency as is shown in Figure 5, whereas the data mining technique detected relatively abnormal sleep onset latency for each participants.

5. Discussions

5.1 Define sleep normality in multiple dimensions

The above results implied that it was not possible to characterize abnormal sleep for all people using a single sleep metric. Anomaly sleep may appear to be normal in one dimension but abnormal in other dimensions. Although sleep has been routinely monitored in multiple dimension in laboratory settings, commercial sleep-tracking technologies tend to simplify sleep quality into a single score such as sleep efficiency or sleep duration. Whereas it is necessary to consider sleep quality in multiple dimensions (e.g. sleep duaration, awakening duration, number of awakenings), our results suggested that not all sleep dimensions were equivalently relevant to everyone. Bed time is the most relevant sleep metric to all participants, whereas sleep onset latency was only related to approximately half of the participants. This compensates the findings of our previous study that individual use different subsets of sleep metrics to define their sleep quality, which in turns was closely related to personal sleep experience [5]. Additionally, findings in clinical sleep research suggest that the relevant sleep metrics may change as a person ages [16].

5.2 Embrace inter-individual differences while acknowledge population patterns

Individuals' sleep anomaly manifested in different sleep dimensions. Even in a same dimension, abnormal sleep patterns of one person could be antithetical to those of another. The results in previous section demonstrated that abnormal sleep could either mean either short or long sleep duration, and either early or late time to go to bed. Difference in sleep patterns between good and poor sleepers is a well-studied topic in sleep research [24, 25]. Inter-individual difference among homogenous good sleepers has not been well explored. Historically, differences in sleep patterns among homogenous population were attributed solely to differences in circumstances or simply personal sleep habits. It was not until recently that inter-individual difference was acknowledged to be biologically determined, at least partially [15]. If genes do play a role, imposing population average (e.g. 8-hour sleep duration or 30-minute sleep onset latency) as a goal to all individuals is questionable, and may even lead to more discontinued and lighter sleep or even chronic insomnia [16]. There is no doubt that individual sleep needs should be taken into account when evaluating sleep quality or detecting sleep anomalies.

It is also worth noting that an absolute individualized sleep evaluation schema without counting in important population-level findings would harm the meaning and correctness of such a schema. Therefore, a natural question is how to understand the sleep needs of each individual and how to strike a balance between such indivdual needs and the population criteron in sleep quality evaluation.

5.3 Limitations of current study

While the results presented in this paper were thought-provoking, this study has the following limitations. First, we did not consider the issue of data quality. Commercial wearable devices have inherent limitations. The collection of self-tracking data was not done in a controlled manner and many human errors could be involved. Therefore, the identified sleep outliers were in fact a mixture of measurement errors, contextual outliers and real sleep outliers. The current approach does not differentiate these cases. Second, we only recruited a small-scale participant cohort in this study. Thus the reliability and generality of the conclusions drawn from the results was limited by the small sample size.

6. Conclusions

In this paper we applied a data mining technique to automatically detect abnormal sleep from personal sleep-tracking data. Defining abnormal sleep as statistical outliers, the data mining task was formulated as multivariate outlier detection problem which counts in the multi-dimensionality of human sleep. Five sleep metrics were selected as features that represent the characteristics of personal sleep. We collected self-tracking sleep data from 7 participants and used the approach to detect abnormal sleep for each of them. We found that population criterion may classify people who were satisfied with their sleep experience as poor sleepers, while fail to detect sleep anomalies for people whose personal sleep distribution deviated from sleep distribution of the population. The detected abnormal sleep patterns demonstrated rich diversities. On the one hand, abnormal sleep manifested in multiple dimensions, and the dimensions were not necessarily equivalent to all people. For another, even in a same dimension, abnormal sleep patterns of one person could be antithetical to those of another. Depending on the person, the "abnormal" could mean both short and long sleep duration, and both early and late time to go to bed. Our study suggested the need for rethinking how consumer devices evaluate sleep quality, taking into acccount interpersonal differences.

Acknowledgments This study was supported by JSPS KAKENHI Grant-in-Aid for Research Activity Start-up (Grant Number 16H07469) and Australia Endeavour Research Fellowship. We would like to thank Prof. James Bailey, Prof. Lars Kulik, Dr. Yuxuan Li and Dr. Kathleen Gray of University of Melbourne for very inspiring and helpful discussions on this work.

Reference

[1] S. Taheri, "The link between short sleep duration and obesity: We should recommend more sleep to prevent obesity," Arch Dis Child, vol. 91, pp. 881–884, 2006.
[2] J. M. Kelly, R. E. Strecker, and M. T. Bianchi, "Recent developments in home sleep-monitoring devices," International Scholarly Research Network, vol. 2012, Article ID 768794, 2012.
[3] A. V. Shelgikar, P. F. Anderson, M. R. Stephens, "Sleep tracking, wearable technology, and opportunities for sleep research and clinical care," Contemporary Reviews in Sleep Medicine, vol.150, no.3, p.732-743, 2016.
[4] T. Ferguson, A. V. Rowlands, T. Olds, and C. Maher, "The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study," International Journal of Behavioral Nutrition and Physical Activity, 12:42, 2015.
[5] Z. Liang, B. Ploderer, W. Liu, et al. "SleepExplorer: a visualization tool to make sense of correlations between personal sleep data and contextual factors," Personal and Ubiquitous Computing, 20:985-1000, 2016.
[6] S. P. Kudyba. Healthcare Informatics: Improving Efficiency and Productivity. CRC Press, 2010.
[7] D. J. Buysse. Sleep health: can we define it? Does it matter? SLEEP, 37(1):9-17, 2014.
[8] W. Liu, B. Ploderer, T. Hoang. In bed with technology: challenges and opportunities for sleep tracking, in proceedings of the Australian computer-human interaction conference (OzCHI 2015).
[9] Z. Liang, B. Ploderer, "Sleep tracking in the real world: a qualitative study into barriers for improving sleep," in proceedings of the Australian computer-human interaction conference (OzCHI 2016). pp.537-541.
[10] M. H. Hall, M. L. Okun, C. W. Atwood, et al. "Measurement of sleep by polysomnography," In Handbook of physiological research methods in health psychology. 2008. Sage Publications; p. 341–67.
[11] Daniel J Buysse, Charles F Reynolds, Timothy H Monk, Susan R Berman and David J Kupfer. The Pittsburgh Sleep Quality Index: A New Instrument for Psychiatric Practice and Research. *Psychiatry research* 28, 2 (1989), 193-213.
[12] R. V. Hogg, E. A.Tanis, and M. J. M. Rao. Probability and statistical inference. McMillan New York, vol. 993, 1977.
[13] P. C. Mahalanobis, "On the generalized distance in statistics", In Proceedings of the National Institute of Sciences of India, vol.2, no.1, pp.49-55, 1936.
[14] P. J. Rousseeuw and K. V. Driessen, "A fast algorithm for the minimum covariance determinant estimator," Technometrics, vol.41, no.3, pp.212-223, 1999.
[15] A. M. Tucker, D. F Dinges, and H. P. A. van Dongen, Trait interindividual differences in the sleep physiology of healthy young adults, J. Sleep. Res, 16, pp.170-180, 2007.
[16] M. M. Ohayon, M. A. Carskadon, C. Guilleminault, M. V. Vitiello, "Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: developing normative sleep values across the human lifespan," SLEEP, vol.27, no.7, 2004.
[17] P. Castillejo, J. F. Martinez, J. R. Molina, and A. Cuerva. Integration of wearable devices in a wireless sensor network for an e-health application. Wireless Communications, IEEE, 20(4), pp.38–49, 2013.
[18] L. Clifton, D. A. Clifton, M. A. F. Pimentel, P. J. Watkinson, and L. Tarassenko. Gaussian processes for personalized e-health monitoring with wearable sensors. IEEE Transaction on Biomedical Engineering, 60(1), pp.193–197, 2013.
[19] A. Smailagic and H. Kenn. New advances in wearable computing. Pervasive Computing, pp.96–99, 2011.
[20] K. L. Knutson, A. M. Ryden, V. A. Mander, and E. van Cauter. Role of sleep duration and quality in the risk and severity of type 2

diabetes mellitus. Arch Intern Med, 166, pp.17681764, 2006.

[21] E. Kasasbeh, D. S. Chi, G. K. Inflammatory aspects of sleep apnea and their cardiovascular consequences. South Med J, 99, pp.58–67, 2006.

[22] S. Stranges, W. Tiqbe, F. X. Gomez-Olive, M. Thorogood, and N. B. Kandala. “Sleep problems: an emerging global epidemic? findings from the indepth who-sage study among more than 40,000 older adults from 8 countries across Africa and Asia,” Sleep 35, pp.1173–1181, 2012.

[23] Y. Kaneita, T. Ohida, Y. Osaki, T. Tanihata, M. Minowa, K. Suzuki, et al. “Insomnia among Japanese adolescents: a nationwide representive survey,” Sleep 29 (12), pp.1543– 1550, 2006.

[24] A. G. Harvey, K. Stinson, K. L. Whitaker, et al, “The subjective meaning of sleep quality: a comparison of individuals with and without insomnia,” Sleep 31(3), pp.383-393, 2008.

[25] K. Adam, M. Tomeny, I. Oswald, “Physiological and psychological differences between good and poor sleepers,” Journal of Psychiatric Research 20 (4), pp.301-316, 1986.

[26] F. P. Cappuccio, L. D’Elia, P. Strazullo, M. A. Miller, “Sleep duration and all-cause mortality: a systematic review and meta-analysis of prospective studies,” Sleep 33 (5), pp.585-592, 2010.