Predicting Depression and Anxiety Mood by Wrist-Worn Sleep Sensor

Shuichi Fukuda*, Yuki Matsuda* Yuri Tani*, Yutaka Arakawa§¶, Keiichi Yasumoto*
* *Nara Institute of Science and Technology*, Nara, Japan,
Email: {fukuda.shuichi.fs5, tani.yuri.tu3, yukimat, yasumoto}@is.nasit.jp.
§ *Kyushu University*, Fukuoka, Japan, Email: arakawa@ait.kyushu-u.ac.jp.
¶ *JST Presto*, Tokyo, Japan.

***Abstract*—In recent years, researches on recognizing daily behavior and psychological / physiological states have been actively conducted to change the behavior of workers working in companies. In this paper, we analyzed Occupational Health questionnaire named DAMS for waking-up time and daily sleep data that are acquired from wearable devices in 2–3 weeks experiment of 60 office workers working in five general companies. By using a machine learning method, our binary Balanced Random Forest model predicts depression, positive, and anxiety moods in two levels, high and low. As a result of Leave One Person Out cross validation, it was confirmed that our model estimated with the F1 values of depression mood: 0.776, positive mood: 0.610, anxiety mood: 0.756. Moreover, we evaluated the variance of the three estimations among subjects by referencing the box chart. As a result, it was confirmed that there is variance in estimation accuracy for each subject.**

***Keywords*—Wrist Sensor, Life log, Machine learning, Wearable computing, Occupational Health**

I. Introduction

Psychological and physiological conditions and environmental conditions have various effects on human behavior and activity. Researchers make efforts to use these data to improve worker's performance and comfort in workplaces. In particular, it is an urgent issue for companies around the world to create workplaces where employees can work actively by grasping and supporting the mental and physical status. To meet those demands, questionnaires are commonly used to grasp the psychological state in the work environment.

In Japan, the revision of the Occupational Safety and Health Act in June 2014 requires that all companies with 50 or more employees have regular stress health checks, which spreads severe management of working hours and stress checks.

In such a background, many studies have been conducted to investigate how the environmental conditions, and psychological and physiological conditions surrounding workers affect worker's performance and comfort.

For example, Annina et al. analyzed questionnaire responses of 154 subjects covering four days each. The questionnaire responses of Pittsburgh Sleep Diary questionnaire [1], which measures sleep quality, and Recovery Experience Questionnaire (REQ) [2], were collected.

As a result of logistic regression analysis, they reported a correlation among a length of working hours per day, a deterioration of sleep quality and REQ recovery status when waking up.

On the other hand, Adam et al. surveyed about work-life balance of 260 people of Long-Haul Truck Driving (LHTD) occupation, an occupation with the greatest health hazard in the United States [3]. A relationship between work-life balance and other factors was investigated by logistic regression analysis and structured analysis. As a result, they report that occupational stress is effective as a predictor of work-life balance decline. The analysis results also show that a decrease in sleep quality and long time of work lead to an increase in occupational stress. From these facts, Adam et al. suggested that managing a good quality of sleep and time will lead to an improvement in the work-life balance of LHTD.

However, Adam et al.'s study has only statistically analyzed the relationship between the work hours answered by the questionnaires. They reported only the correlation acquired from statistical analysis. It is not possible to monitor the "daily" mental and physical condition of each worker, to provide individual support, and to promote behavior change.

On the other hand, there are many studies aimed to improve worker's performance, comfort and promoting behavioral change [4], [5]. In order to build a system that leads to better behavior, it is essential that the system automatically recognizes daily behavior patterns, mental and physical states.

For instance, Amemori et al. proposed an estimation method for HRQOL (Health-Related Quality of Life) using smart devices [6]. In their method, life logs and QOL questionnaires from a wearable device and a smartphone were collected for one subject for 150 days. Next, HRQOL is estimated by machine learning model with features of wearable devices.

However, their method uses dataset obtained from only one student subject, the generality between subjects has not been verified, and there is a problem that it lacks versatility. Besides, they use the Empatica E4 wristband as a wearable device, although it costs about $2,000 and raises the cost per person.

From these backgrounds, the purpose of our study is to automatically grasp the mental and physical status of office workers working in companies. Moreover, we aim to support them to prompt behavioral transformation and improve worker's performance and comfort. By collecting daily life-log data of 60 workers, which are an occupational health questionnaire (named DAMS) and daily sleep data that can be acquired from wearable devices and smartphones, we built a binary machine learning model to predict a depression, positive and anxiety moods calculated from DAMS questionnaire.

As a result of Leave One Person Out cross validation, we confirmed that the occupational health index can be estimated with F1 value of depression mood: 0.776, positive mood: 0.610, anxiety mood: 0.756. Also, each confusion matrix showed that depression and anxiety moods were generally well estimated. But, the low level of positive mood could not be estimated well.

As a result of three box diagrams that show a variance in the estimation of each mood between subjects, it was confirmed that the variance in depression estimation was small and a width of the first and third quartiles was 0.24. It means that, according to the variance among subjects with a figure of 0.24, some subjects may be estimated with high accuracy, but in some cases, some other subjects may be estimated with low accuracy.

By creating a general-purpose model among 60 subjects, it will be possible for office workers in the future to automatically recognize their daily mental and physical states by wearing a wearable device, and to support them accordingly.

II. Related Work

A. Statistical Analysis Research on Sleep

Annina et al. analyzed a relation between sleep, recovery experience and working hours for 154 subjects for four days each by using logistic regression [2]. Their questionnaires are Pittsburgh Sleep Diary questionnaire, which measures the quality of sleep and Recovery Experience Questionnaire (REQ). As a result of logistic regression analysis, they reported a correlation between daily work hours, poor sleep quality and worsening REQ experience during awake.

Adam et al. surveyed with questionnaires about occupational stress, sleep quality, and work-life balance in 260 people in an occupation of Long-Haul Truck Driving (LHTD), the occupation with the highest health hazards in the United States [3]. As a result of logistic regression analysis and structured analysis, it has been reported that occupational stress is effective as a predictor of a decline in work-life balance. Their analyses also show that decreases in sleep quality and time lead to an increase in occupational stress. From these facts, Adam et al. suggested that securing a good quality of sleep and time will lead to an improvement in the work-life balance of LHTD.

Hülsheger et al. conducted a fatigue assessment four times a day over five working days for 133 employees, and investigated a relationship between changes in fatigue level and sleep quality [7]. As a result of growth curve analysis, fatigue was found to decrease on average in the morning, reach the lowest point around noon, and then increase until bedtime. As a result of analyzing a relationship between daily fatigue patterns and sleep quality, it was confirmed that when the quality of sleep is low, next day's fatigue will be high at waking up and decrease at noon and then increase until going to bed. When sleep quality is high, fatigue will be low and stable until noon and increase between the end of work and bedtime.

Hülsheger suggested that the transition of the fatigue level in a day depends on the quality of sleep. If the fatigue transition pattern can be predicted, it is thought that employees can gain support and better behavioral change based on the predicted transition of it.

B. Questionnaire Analysis by Machine Learning

Sano et al. suggested a machine learning method which estimates student's academic ability (GPA), sleep quality index (Pittsburgh Sleep Quality Index), and mental health composite score obtained from a questionnaire [8]. Using the questionnaires responses for total 1980 days (for 30 days for each of 66 student subjects), they extracted 700 features that are acquired from the smartphones and wearable sensors. Then they built a binary SVM model. As a result of Leave One Person Out cross validation, it has been reported they achieve an accuracy exceeding 80%.

In Sano et al.'s study, targeted subjects are only students, so it is impossible to grasp the mental and physical status of office workers working in company.

Amemori et al. proposed a simple estimation method of HRQOL using smart devices [6]. In this method, life logs from smartphones and wearable devices, and ground truth data from HRQOL questionnaires are collected for 150 days. Then they extracted features from wearable sensor data. Furthermore, they constructed a HRQOL estimation model by using a machine learning method. In the initial stage of learning, the estimation model is constructed using all the features, but the model is reconstructed every day by using newer measured data. In addition, the model is optimized for each individual.

it was confirmed that the ground truth could be followed with a correlation coefficient of 0.646. Since this method uses only smartphones and wristband type wearable devices for estimation, it is possible to easily evaluate HRQOL without restricting the movement of subjects in daily life.

However, their method uses only data obtained from a single student subject. Its generality has not been verified, and there is a problem that it lacks versatility. Besides, the Empatica E4 wristband which unit price is about $2,000 is used. Therefore, it raises the cost per person.

C. Novelty of Our Research

In the above-mentioned statistical analysis, only the statistical correlation and the relationship are reported by analyzing sleep quality, recovery experience, fatigue and stress obtained from the questionnaires. Therefore, it is impossible to recognize the mental and physical *daily* condition of each worker working for a company. In a machine learning questionnaire analysis, it is pointed out that it is not targeted at office workers working in companies, and there are few subjects and poor generalization.

In this paper, in order to grasp the mental and physical state of office workers, we proposed an occupational health index estimation method using a reasonable wearable wristband. We collected occupational health questions (named DAMS [9]) responses of waking-up time in 2–3 weeks experiment including

60 office workers working in five Japanese companies. Next, we built a machine learning model which estimates two level of occupational health index (depression, positive, and anxiety moods). Finally, we evaluated its F1 values and variance of accuracy among 60 subjects.

III. Proposed Method

A. Data Collection

The dataset to construct the machine learning model was collected from 2–3 week experiment including 60 office workers working in five Japanese companies.

For appropriate data collection, a six days practice period was set up before the data measurement period. The preliminary questionnaire on static characteristics such as gender and age of the subjects was answered too. During the experiment, subjects lived with a Fitbit wearable sensor for acquiring sleep status [10] throughout the weeks, and at the same time, they answered questionnaires about occupational health on their smartphones.

B. Data Cleaning of Questionnaire Responses

DAMS (Depression and Anxiety Mood Scale) [9] was used as a occupational health index in this study. DAMS is a questionnaire to measure three degrees of depression, positive, and anxiety moods. In order to measure those moods, the words in questions expressing mood such as *lively*, *dark*, and *worried*, can be selected from seven levels of how well you feel. The method of calculating the depression mood score is the sum of a seven-point evaluation of *dark*, *sunk*, and *disgusting* question items. Thus the score can take a value in the range of 0–18. Similarly, positive mood scores are *fund*, *joyful*, and *lively*. Also, anxiety mood are *uneasy*, *disturbing* and *worried*.

A labeling method of ground truth for machine learning is the same as Sano et al., whose research estimated the mental index obtained from the questionnaires [8]. Their method of creating ground truth rates the upper 20% and lower 20% of the mental index score as high and low, respectively. Middle 60% data are discarded because it contains a lot of ambiguity of questionnaire responses.

In this study, similarly, the upper 20% and lower 20% of the depression, positive, and anxiety scores obtained from the DAMS questionnaire were set to two values, high and low. The score distribution and the range of upper and lower 20% for depression, positive, and anxiety mood scores are shown in Figs. 1, 2, and 3.

The upper and lower 20% of depression, positive, and anxiety scores are as follows: Depression mood (Top 20% score is nine points, Bottom 20% score is zero points), Positive mood (Top 20% score is nine points , Bottom 20% score is three points), anxiety feeling (Top 20% score is twelve points, Bottom 20% score is three points).

The total number of DAMS questionnaire responses was 840, and it was decreased to 685 in a process of removement such as forgetting to wear the fitbit and being disable to concat to questionnaires dataset. Lastly, the total number of ground truth data after labeling was 274.

TABLE I: List of Features

	Feature Names
1	Total Sleep Minutes
2	Total Wake Minutes
3	Light Sleep Minutes
4	Rem Sleep Minutes
5	Deep Sleep Minutes
6	Total Wake Time Ratio
7	Light Sleep Time Ratio
8	Rem Sleep Time Ratio
9	Deep Sleep Time Ratio
10	Number of Wake
11	Number of Light Sleep
12	Number of Rem Sleep
13	Number of Deep Sleep

C. Features Extraction

Table I shows a list of features related to sleep, which is an input to the model for estimating occupational health indices.

In this study, the sleep status of the subjects was obtained from Fitbit Charge 3 [11]. Fitbit Charge 3 can measure not only wake-up time and bedtime detected automatically, but also four levels of sleep, such as REM sleep, and deep sleep during one night. Also, these four levels of sleep time and their ratio were added into features.

The sleep features are extracted from the sleep of the day immediately before answering the DAMS questionnaire when waking up. There are 13 features in total.

D. Model building

We built a BRF (Balanced Random Forest) model for estimating occupational health indices [12], [13]. Table II shows the parameters of the BRF model for the estimation of depression, positive, and anxiety moods that were adjusted by a grid search method.

IV. Results & Discussion

A. Evaluation Method

By using sleep features and two levels of the depression, positive, and anxiety scores as ground truth, we evaluated our BRF model with Leave One Person Out cross validation method.

The Leave One Person Out cross validation method divides the dataset to each subject data. Next, the method chooses to evaluate data for only one subject from the dataset, and builds a model by using all other subjects' data as a learning data set. It is a method to evaluate a generalization performance among subjects in the model by repeating it for each subject. Accuracy, Precision, Recall, and F1 Value were used as evaluation metrics.

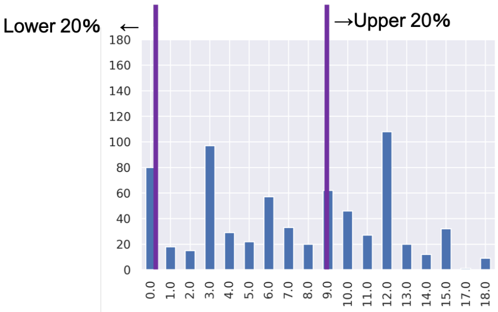

Fig. 1: Distribution of Depression Score

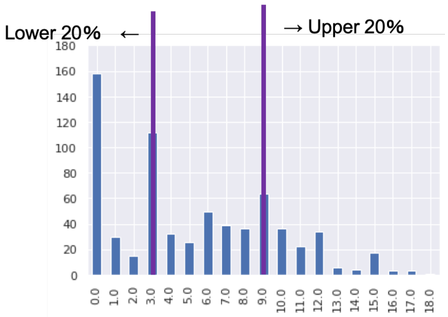

Fig. 2: Distribution of Positive Score

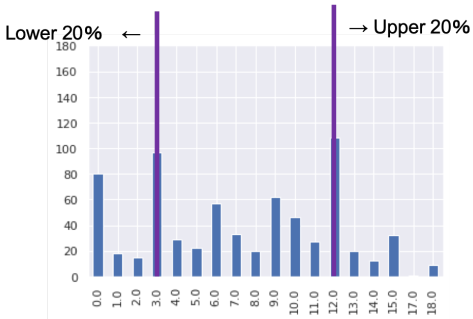

Fig. 3: Distribution of Anxiety Score

TABLE II: Parameters of BRF

	Depression	Positive	Anxiety
bootstrap	True	True	True
class_weight	balanced	balanced	balanced
criterion	gini	gini	gini
max_depth	None	3	3
max_features	3	3	5
max_leaf_nodes	1	1	1
min_samples_leaf	2	2	2
min_samples_split	2	2	2
min_weight _fraction_leaf	0	0	0
n_estimators	100	100	30
n_jobs	-1	-1	-1
oob_score	False	False	False
random_state	1	1	1
sampling_strategy	all	all	all
verbose	0	0	0

B. *Results of Leave One Person Out Cross Validation*

Table III shows the results of the cross validation among subjects. The F1 Value of the depression mood was 0.732, the highest among three. In addition, the variance of Accuracy, Precision, Recall, and F1 Value among subjects are shown in Figs. 4, 5, and 6.

In the labeling method proposed by Sano et al., when only the top/bottom 20% response results were extracted, the data for 60 people and two weeks decreased to 40%. Moreover, the number of data samples of each subject included in the extracted data varies. As a result, the average value of F1 Value in Leave One Person Out cross validation includes the results of subjects with small number of data samples.

Therefore, Table IV shows the results of recalculating Accuracy, Precision, Recall, and F1 Value, excluding data of subjects whose number of data is less than one week. After the data labeling, there were 42, 44, and 42 subjects in the depression, positive, and anxiety moods responses respectively. Then, the number of subjects after excluding subjects with less than 1 week of data was 20, 19, and 21. From Table IV, F1 value of depression was the highest, which

TABLE III: Results of Leave One Person Out Cross Validation

	Depression	Positive	Anxiety
Accuracy	0.639	0.640	0.565
Precision	0.917	0.902	0.704
Recall	0.639	0.640	0.565
F1 Value	0.732	0.696	0.600

TABLE IV: Accuracy of Subjects with Data for one week or more

	Depression	Positive	Anxiety
F1 Value	0.776	0.610	0.756

was 0.776. Whereas, F1 value of positive mood was poor with the value of 0.610.

Regarding a variance in three estimations among subjects, it can be confirmed that there is a variance between subjects from Figs. 4, 5, and 6.

The variance in depression among the subjects was the smallest, and the width of the first and third quartiles was 0.24 with F1 value. On the contrary, the variance in estimation of positive and anxiety was 0.53 and 0.57 with the first and third quartiles.

Thus, according to the variance among subjects of 0.24, some subjects' moods may be estimated with high accuracy, but in some cases, some other subjects' moods may be estimated with low accuracy. It is considered that the estimation accuracy may be improved by giving learning data to each user based on our model and personalizing it.

C. *Overview of Estimation*

Confusion matrices of three estimation results for depression mood, positive mood, and anxiety mood are shown in Figs. 7, 8, and 9.

From those results by referring confusion matrixes, we were able to estimate depression and anxiety moods well in two categories. However, our model mispredicted in the positive mood estimation with Fig. 8. This is considered that it was difficult to answer the positive word question items such as *fun*, *joyful*, and *lively* when people are awake.

Tables V, VI, and VII show importance of the overall features in depression mood, positive mood, and anxiety mood

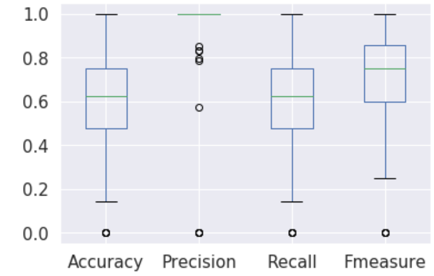

Fig. 4: Box plot of Depression Mood

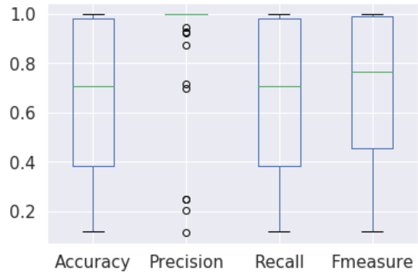

Fig. 5: Box plot of Positive Mood

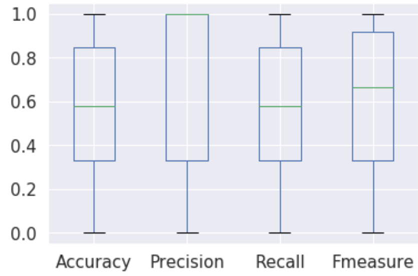

Fig. 6: Box plot of Anxiety Mood

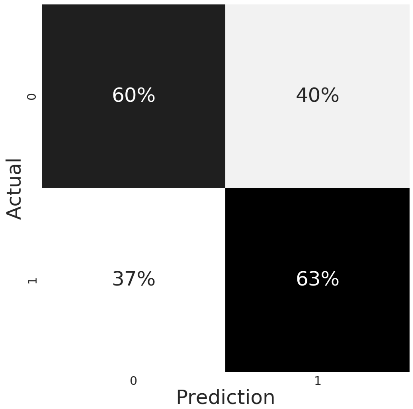

Fig. 7: Confusion Matrix of Depression Mood

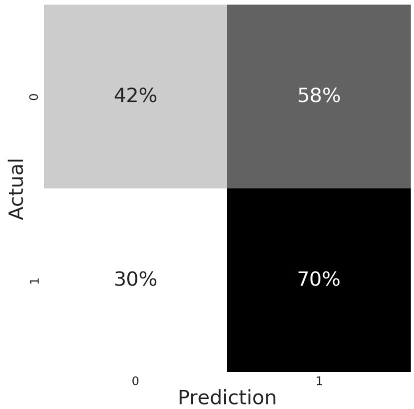

Fig. 8: Confusion Matrix of Positive Mood

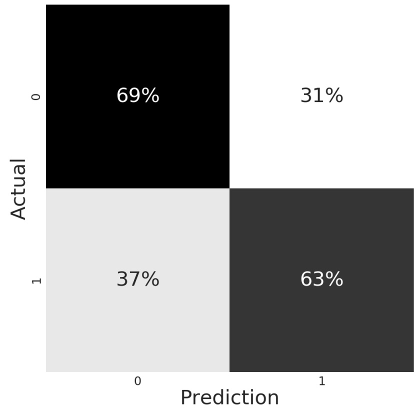

Fig. 9: Confusion Matrix of Anxiety Mood

TABLE V: Features Importance of Depression Mood Estimation

	Features	Importance
1	Light Sleep Time Ratio	0.146339
2	Rem Sleep Minutes	0.117653
3	Rem Sleep Time Ratio	0.117653
4	Number of Rem Sleep	0.098994
5	Total Wake Minutes	0.071501
6	Deep Sleep Time Ratio	0.070383
7	Deep Sleep Minutes	0.068154
8	Light Sleep Minutes	0.066242
9	Total Sleep Minutes	0.065598
10	Total Wake Time Ratio	0.063514
11	Number of Wake	0.049613
12	Number of Light Sleep	0.044701
13	Number of Deep Sleep	0.026987

TABLE VI: Features Importance of Positive Mood Estimation

	Features	Importance
1	Light Sleep Time Ratio	0.125438
2	Rem Sleep Time Ratio	0.124797
3	Rem Sleep Minutes	0.116199
4	Total Sleep Minutes	0.098601
5	Deep Sleep Minutes	0.088327
6	Light Sleep Minutes	0.084920
7	Number of Rem Sleep	0.065996
8	Total Wake time Ratio	0.064743
9	Total Wake Minutes	0.056930
10	Deep Sleep time Ratio	0.053145
11	Number of Wake	0.046709
12	Number of Deep Sleep	0.040319
13	Number of Light Sleep	0.033874

estimation. In each estimation, it was confirmed that the top five items of the importance include a characteristic related to REM sleep. Similarly, a feature of the light sleep time ratio is also in the top five. Based on these facts, it is considered that the features of the time ratio of REM sleep and light sleep are effective in estimating depression, positive, and anxiety moods.

V. Conclusion

In this paper, by collecting DAMS questionnaire responses at waking-up time and daily sleep data that are acquired from wearable devices in the course of 2–3 weeks experiment of 60 office workers working in five general companies, We constructed a Balanced Random Forest model that predicts depression, positive, and anxiety moods in two categories, which are high and low. As a result of Leave One Person Out

TABLE VII: Features Importance of Anxiety Mood Estimation

	Features	Importance
1	Rem Sleep Time Ratio	0.109272
2	Rem Sleep Minutes	0.100533
3	Light Sleep Time Ratio	0.094728
4	Deep Sleep Time Ratio	0.083043
5	Total Wake Time Ratio	0.082017
6	Total Sleep Minutes	0.077754
7	Light Sleep Minutes	0.077330
8	Number of Wake	0.074166
9	Deep Sleep Minutes	0.067463
10	Number of Rem Sleep	0.066891
11	Number of Light Sleep	0.062715
12	Total Wake Minutes	0.062314
13	Number of Deep Sleep	0.041775

cross validation, it was confirmed that our model can estimate with F1 value of depression mood: 0.776, positive mood: 0.610, anxiety mood: 0.756. The confusion matrixes showed that the low level of positive mood could not be estimated well, but the depression and anxiety moods could be estimated well overall. As a result of three box diagrams that show the variance in the estimation of each mood between subjects, it was confirmed that the variance in depression estimation was small and the width of the first and third quartiles was 0.24 with F1 value. It means that, according to the variance among the subjects with a figure of 0.24, some subjects may be estimated with a high accuracy, but in some cases, some other subjects may be estimated with a low accuracy.

Future tasks include adding features such as age and gender. Also, we will compare the accuracy of other machine learning models to improve the accuracy. In addition, for practical use, when personalizing to subject based on this model, we will confirm how much accuracy will be improved by giving the subject's data.

ACKNOWLEDGMENTS

This research is partially supported by JST PRESTO, and IoT-NA group, the research community of CAN (Consortium for Applied Neuroscience).

REFERENCES

[1] T. H. Monk, C. F. REYNOLDS III, D. J. Kupfer, D. J. Buysse, P. A. Coble, A. J. Hayes, M. A. Machen, S. R. Petrie, and A. M. Ritenour, "The pittsburgh sleep diary," *Journal of sleep research*, vol. 3, no. 2, pp. 111–120, 1994.

[2] A. Ropponen, M. Härmä, B. Bergbom, J. Nätti, and M. Sallinen, "The vicious circle of working hours, sleep, and recovery in expert work," *International journal of environmental research and public health*, vol. 15, no. 7, p. 1361, 2018.

[3] A. Hege, M. K. Lemke, Y. Apostolopoulos, B. Whitaker, and S. Sönmez, "Work-life conflict among us long-haul truck drivers: Influences of work organization, perceived job stress, sleep, and organizational support," *International journal of environmental research and public health*, vol. 16, no. 6, p. 984, 2019.

[4] Y. Arakawa, "[invited paper] sensing and changing human behavior for workplace wellness," *Journal of Information Processing*, vol. 27, pp. 614–623, 2019. [Online]. Available: https://www.jstage.jst.go.jp/article/ipsjjip/27/0/27_614/_article/-char/en

[5] Z. Zhang, Y. Takahashi, M. Fujimoto, Y. Arakawa, and K. Yasumoto, "Investigating effects of interactive signage–based stimulation for promoting behavior change," *Computational Intelligence*, vol. 35, no. 3, pp. 643–668, 2019. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/coin.12234

[6] C. Amenomori, T. Mizumoto, H. Suwa, Y. Arakawa, and K. Yasumoto, "A method for simplified hrqol measurement by smart devices," in *Wireless Mobile Communication and Healthcare*, 2018.

[7] U. R. Hülsheger, "From dawn till dusk: Shedding light on the recovery process by investigating daily change patterns in fatigue." *Journal of Applied Psychology*, vol. 101, no. 6, p. 905, 2016.

[8] A. Sano, A. J. Phillips, Z. Y. Amy, A. W. McHill, S. Taylor, N. Jaques, C. A. Czeisler, E. B. Klerman, and R. W. Picard, "Recognizing academic performance, sleep quality, stress level, and mental health using personality traits, wearable sensors and mobile phones," in *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*. IEEE, 2015, pp. 1–6.

[9] L. A. Clark, "The anxiety and depressive disorders: Descriptive psychopathology and differential diagnosis." 1989.

[10] M. de Zambotti, A. Goldstone, S. Claudatos, I. M. Colrain, and F. C. Baker, "A validation study of fitbit charge 2™ compared with polysomnography in adults," *Chronobiology international*, vol. 35, no. 4, pp. 465–476, 2018.

[11] "Fitbit Inc. : Fitbit charge 3," available from https://www.fitbit.com/jp/charge3 (accessed 2019-10-20).

[12] M. Khalilia, S. Chakraborty, and M. Popescu, "Predicting disease risks from highly imbalanced data using random forest," *BMC medical informatics and decision making*, vol. 11, no. 1, p. 51, 2011.

[13] Y. Xie, X. Li, E. Ngai, and W. Ying, "Customer churn prediction using improved balanced random forests," *Expert Systems with Applications*, vol. 36, no. 3, pp. 5445–5449, 2009.