Toward the Analysis of Office Workers' Mental Indicators Based on Wearable, Work Activity, and Weather Data

Yusuke Nishimura, Tahera Hossain, Akane Sano, Shota Isomura, Yutaka Arakawa, Sozo Inoue

Abstract In recent years, many organizations have prioritized efforts to detect and treat mental health issues. In particular, office workers are affected by many stressors, and physical and mental exhaustion, which is also a social problem. To improve the psychological situation in the workplace, we need to clarify the cause. In this paper, we conducted a 14-day experiment to collect wrist-band sensor data as well as behavioral and psychological questionnaire data from about 100 office workers. We developed machine learning models to predict psychological indexes using the data. In addition, we analyzed the correlation between behavior (work content and work environment) and psychological state of office workers to reveal the relationship between their work content, work environment, and behavior. As a result, we showed that multiple psychological indicators of office workers can be predicted with more than 80% accuracy using wearable sensors, behavioral data, and weather data. Furthermore, we found that in the working environment, the time spent in 'web conferencing','working at home (living room)', and 'break time (work time)' had a significant effect on the psychological state of office workers.

Yusuke Nishimura
Kyushu Institute of Technology, Fukuoka, Japan, e-mail: `nishimura@sozolab.jp`

Tahera Hossain
Kyushu Institute of Technology, Fukuoka, Japan, e-mail: `tahera@sozolab.jp`

Akane Sano
Rice University, Houston, Texas, USA, e-mail: `akane.sano@rice.edu`

Shota Isomura
NTT Data Institute of Management Consulting, Inc., Tokyo, Japan, e-mail: `isomuras@nttdata-strategy.com`

Yutaka Arakawa
Kyushu University, Fukuoka, Japan, e-mail: `arakawa@ait.kyushu-u.ac.jp`

Sozo Inoue
Kyushu Institute of Technology, Fukuoka, Japan, e-mail: `sozo@sozolab.jp`

1 Introduction

In recent years, research on mental health and well-being has attracted a lot of attention as a way to improve the quality of one's personal and professional life [1, 2]. Smart sensing, machine learning, and data analytics have been rapidly utilized to get new insights into human health, lifestyle, personality, and other human characteristics and behaviors.

It has been studied that many physical and mental disorders appear in a variety of physiological and behavioral manifestations before being diagnosed [3, 4]. If everyday health and well-being can be monitored using ubiquitous sensors, this assessment could help users reflect on their behaviors to prevent severe mental and physical disorders and help clinicians monitor users' conditions and diagnose disorder. In our daily life, work occupies a major part of individuals' days. It is necessary for employees and individuals to design a workplace that impacts them positively in order to enjoy a certain quality of life. Workplace stress, anxiety, and depression are harmful to human health and productivity, with major financial consequences. The number of companies that employ stress assessments to ensure a high-quality work environment has risen by 12% in 2021 compared to 2012 [5]. In addition, the number of companies that use the assessment results has also increased.

There are several reasons why companies should work to improve their employees' mental health. The first reason is the employee retention rate. The past research has shown that work environments with low levels of job satisfaction have higher turnover rates than those with high levels [6, 7]. The second reason is the maintenance of employee health. When people are under high negative psychological stress in the workplace, they are increasing the risk of physical illness as well as mental illness [8]. Well-being is also expected to have a significant impact on work productivity, and a high-wellbeing level group of individuals tends to perform better than a low-wellbeing level group of individuals [9, 10]. The majority of these studies rely on self-reported assessments and sensor data acquired passively from smartphones and other wearable devices.

In this research, we collected real-world 2-week data from N=100 office workers engaged in intellectual work to discover knowledge about the impact of their behavior on the psychological measures. The dataset collected in this study consisted of three main types. (1) daily activity data using a Fitbit smartwatch which the subjects continuously wore during the experiment, (2) psychological data from questionnaires. Subjects self-reported their psychological states using their smartphones three times a day, (3) behavioral data self-reported on the smartphone application to record their work activity and environment-related information throughout the experiment. The behavioral data includes the type of work tasks as well as the detailed situation when and how the work tasks were performed, such as "with whom", "the type of the work", and "the place where the work was performed". Also, this experiment was conducted while remote work was introduced. Therefore, the data were collected under the condition of working both in the office and at home.

We analyzed the collected data in two ways: (1) we developed prediction models for six psychological indicators of office workers each night and the following morning, and (2) we analyzed the contribution of features and SHAP (SHapley Additive exPlanations) value in the prediction models to analyze the correlation between behavioral data and psychological data. Our results showed that multiple psychological indicators of office workers can be predicted with an accuracy of 80% or more using wearable sensors, behavioral data, and weather data. Furthermore, we found work related factors that affected the psychological indicators of office workers. For example, amoung behavioral data, the time spent on "eating," "working alone," "hobbies," "resting," and "traveling" had high effects on the psychological situation. In addition, "with whom the work was done" and "whether the work was standardized" were found to have high effects.

The main contributions of this paper are:

- The real-world office worker multimodal dataset, where the dataset contains objective physiological and behavioral sensor data, and work activities and work environment information with different psychological state information. The dataset depicts the status of real office workers states performing their everyday work under real-life stressors.
- Prediction of different mental states of office workers.
- Correlation analysis between behavior and psychological state of office workers to reveal the factors in the workplace that affect their psychological state.

We expect this dataset and analyses will contribute to design work behavior and environment and eventually improve the performance of employees, the turnover rate, and the solution of the shortage of human resources in the workplace. The remaining part of the paper proceeds as follows. First, section 2, begins by reviewing the related literature on estimating stress, mood, and mental health. Section 3 explains data collection and explanation of data attributes. Section 4 describes the feature extraction of multimodal data, pre-processing, and model building process. Section 5 presents results about prediction models and correlation analysis among different psychological indexes and the behavior of office workers. In Section 6, results are discussed and the conclusions drawn are presented with some future work points in Section 7.

2 Related Research

Mood, health, and stress are three widely investigated wellbeing labels. In recent years, research on estimating stress, mood, and mental health indicators using sensor data has become popular [11, 12, 13] . Koldijk et al. [14] used a multimodal set of sensor data including computer logs, facial expressions, and posture data using webcam and Kinect, and to detect stress with 90% accuracy. Alberdi et al. predicted stress and workload from ecological data and behavioral data of office workers under real stress factors in smart office environments [15].

Sano et al. [16] used smartphones, wearable sensors, and survey data to predict student performance, stress, and mental state with a precision of 67 to 92%. In another study, Li et al. [1] aim to create a deep learning framework that can predict human well-being from raw sensor signals to assist individualized long-term health monitoring and early-warning systems. The proposed framework was tested over a period of 6391 days utilizing wearable sensor data and wellbeing labels obtained from college students. Robles et al. [17] used log data from smartphones, wearable sensors, and social apps to develop a framework that can predict stress. Studies by Amamori and others estimated survey items related to HRQOL (Health-Related Quality of Life) using sensor data from wristband terminals and positional data from smartphones [18].

On the other hand, many researchers are working to discover workers' physical and mental discomfort experienced as a result of their day-to-day activities, workload, and work environment, which may lead to decreased work performance. Improving employee wellness services for managing these issues is critical, but understanding who requires attention is a prerequisite step toward proactive care. In this regard, Feng et al. [19] examined the impact of irregular work hours on health and showed the significant negative impact on health. According to the findings, night shift nurses were more sedentary and showed lower levels of life satisfaction than day shift nurses. There are many other studies on the psychological analysis of workers. Lee et al. [20] used intelligent worker physiological sensor data to predict the state of concentration at work. Fukuda et al. [21] used wristband wearable sensors to collect sleep data and estimated mood indicators from questionnaire surveys. The classification accuracy was 60-73%, which indicated the contribution of psychological prediction in sleep. In another worker stress study, the accuracy of psychological indicators has reached 71% [22], where, they used smartphones accelerometer sensor data to monitor behaviour of subjects with their stress level from self-assessment questionnaire data.

Alexandros et al. [23] used data collected from banded wearable devices and smartphones to predict the mood of workers in an office environment. In this study, they used five levels of prediction for eight different types of moods, and the results were 70.6% for personalized prediction and 62.1% for generalized prediction using the Bagged Ensembles of Decision Trees. Mirjafari et al. [24] differentiated *'Higher'* and *'Lower'* job performers in the workplace. They used mobile sensing, wearable, and beacon data to train a gradient boosting classifier that can classify workers as higher or lower performers. The majority of these studies have relied on self-reported assessments as well as sensor data acquired passively from smartphones and other wearable devices. Table 1, shows the summary information about prior studies that predicted wellness and psychological state using sensor module and other data. .

It has been shown that in many early researches passively collected sensor data, biometric sensor data, images, self-reported data and acceleration data often aim to improve stress and mood recognition of office workers and students [27, 28], but few studies have used the type of person's behavior as an explanatory variable [15][23][24][17][18][29]. There are also many studies that estimate mood and stress, but few studies have estimated fatigue, productivity, work engagement, and work self-evaluation, etc., which are important for the performance of office workers by the same explanatory variable. Although several studies have been conducted to estimate psychological indicators based on data from working, there are few cases

Table 1: A summary of previous studies for predicting wellness and psychological state

Research Target	Participants (N)	Sensors/Module	Data Used
Understanding how daily behaviors and social networks influence self-reported stress, mood, and other health or well-being-related factors [25]	University Students, N=201	Wrist-based sensors electronic diaries (e-diaries)	Questionnaires data acceleration and ambient light data.
Differentiating higher and lower job performers in the workplace [24]	Working professionals, N=554	Smartphones (i.e., Android and iOS), wearables (i.e., Garmin vivosmart) and bluetooth beacon	Mobile sensing and daily survey data.
Forecasting Personalized Mood, Health, and Stress [1]	College students, N=239	Wearable sensor and self-report assessments	Skin temperature, skin conductance, and acceleration; self-reported mood, health and stress scored
Multimodal analysis of physical activity, sleep, and work shift in nurses with wearable sensor data [19]	Nurses in a large hospital, N=113	Fitbit, self-reported assessments of affect and life satisfaction	Sleep pattern data from Fitbit, Demographic data
Detecting affective flow states of knowledge workers using physiological sensors [20]	Industrial research lab professionals, N=12	Physiological sensors, webcam-based techniques for measuring a worker's pulse, respiration, affect, or alertness.	Sensors and webcam data collected in a controlled lab setting
Using smart offices to predict occupational stress [15]	Office workers, N=25	Mobi (TMSI) sensors with self-adhesive electrodes for ECG and Skin Conductance Level (SCL)	Heart Rate (HR), heart Rate Variability (HRV), SCL, self-reported stress and mental workload scores
Forecasting depressed mood based on self-reported histories via recurrent neural networks [26]	Random participants using the application from apps store, N=2382	Smartphone application called 'Utsureko' for collecting data from users	Self-reported historical data of mood, behavior log and sleeping log
Forecasting stress, mood, and health from daytime physiology in office workers and students [27]	Employees at a high-tech company in Japan and college students, N=240	Skin conductance, skin temperature, and acceleration from a wrist sensor	Self-reported data and wrist sensor data.
Understanding a relationship between stress and individual's work role [28]	Volunteers working in a research division of a large corporation, N=40	Physiological data was collected from a heart rate monitor worn around the chest and a FitBit.	Cardiovascular data, multiple daily self-reports of momentary affect, and filled out a one-time assessment of the global perceived stress data.
Mood recognition at work using smartphones and wearable sensors [23]	Recruited 4 users (researchers) to take part in this study which was conducted in an office environment, N=4	Toshiba Silmee wristband sensor and smartphone app 'HealthyOffice' to collect self-reporting data	Heart rate, pulse rate, skin temperature, 3-axial acceleration and self-reported data

where people working for companies are tested as subjects, and it was difficult to collect data in actual workplaces. There are few studies that have focused on the analysis of the work environment during work behavior, and there have been few analyses that take into account where workers work, the people they work with, and the work environment.

Based on the above, this paper aims to analyze the psychological state of office workers to reveal the relationship between their psychological states and their work content, work environment, and behavior by using sensor data, self-reported time-series work task/environment data, and psychological index data together. The data collected in this research depicts the behaviors of real office workers performing their natural office work under real-life stressors. We believe that this will contribute to the development of tools for improving the occupational health of office workers. This study will be also expected to help find knowledge to improve working methods by revealing the relationship between psychological indicators and behavior.

3 Data Overview

This experiment was led by NTT Data Management Institute and conducted to collect data from several companies that cooperated with the project. Sixty-three male and 37 female workers with an average age of 42.1 years participated in a 14-days data collection experiment in January 2021. The data collected in this experiment included sensor data, self-reported work task/environent data, weather data, and psychological index data obtained from questionnaires.

3.1 Sensor data

The sensor data were collected using a wrist-band sensor, Fitbit [30]. Fitbit data included data on calorie consumption, heart rate, sleep characteristics, metabolic equivalent, step count, floor count, and activity characteristics. Heart rate was measured every 5 seconds, and all other Fitbit data were measured every minute.

3.2 Work task/environment data

In order to collect behavioral data and questionnaire data, we used a nursing care behavior recording application fonlog [31] developed by Inoue and others for office workers. Behavioral data were collected by labeling the participants' behavior of what kind and what time they were doing. Participants provided an average of 9.1 behavioral labels per day. Table2, shows the questionnaire items related to work. The types of behavioral labels are as follows: face-to-face meetings, meals, single work, hobby/break, housework/child-rearing, rest (during business hours), travel, web meetings, collaboration (with the communication), telephone (meetings), non-business work. In particular, activities in the work were recorded not only by entering the type but also details of the task and the environment in which the task was performed also recorded.

Table 2: Types of records and their values related to task description and work environment.

Activities	Records	Values
Work Tasks	Task Description	Planning Task Development Task Sales Task Management Task Field Task Office Task
	Scope of the work	Core Task Non-core Task
	Novelty of the work	Standardized Task Non-standardized Task
	Position in the work	Managers Operators and Paticipants Collaborators Managers and Operators
Work Environment	Work Environment	Home (place for work) Home (living) Home (other) Workspace(outside) Store and Outside Work Place
	Task Situation	Alone With Others (no interaction) With Others (colleagues) With Others (family)
	Work Environment Assessment	Very Comfortable Comfortable Neither Uncomfortable Extremely Uncomfortable

3.3 Psychological measures

In the questionnaire on psychological conditions, six indicators were used to evaluate the results.

- Depression and Anxiety Mood Scale (DAMS)[32]
  DAMS questions to evaluate each strength in 7 stages with respect to positive mood, depressive mood, and anxiety mood (9 questions). Therefore, each score is a value between 0 and 6. Each of the DAMS scores has been shown to be highly

suitable for the study of cognitive-behavioral models that include both depression and anxiety, as they can sensitively capture changes in mood. Positive mood is the average of degree of "vivid", "happy", and "fun". Depressive mood is average of the degree of "gloomy", "Unpleasant", and "Sinking". Anxiety mood is average of the degree of "worried," "anxious," and "concerned".

- Subjective Pain [33]
  Subjective pain is a questionnaire that evaluates the type and degree of fatigue an individual is aware of. There are five questions each about sleepiness, discomfort, and lethargy, and each item is rated on a five-point scale, and the average of these scores is used as the score. Each score is a value between 0 and 4 (15 questions).
- Recovery Experience [34]
  Recovery experience is a questionnaire to evaluate how an individual recovers from work in leisure time. The average score for each question is the average score for each question in the seven-step evaluation for "psychological distance to work", "relaxed", "learned new things”, and "decided what to do by myself".Each score is a value between 0 and 6 (4 questions).
- Work Engagement [35]
  Work engagement questions assess how enthusiastic an individual is about their work. One question each on "vitality", "enthusiasm" and "immersion" in work and the average of the seven steps is used as the work engagement score Each score is a value between 0 and 6 (3 questions).
- Productivity [36]
  Health and labor performance questions designed by the World Health Organization. The score is the self-assessment of the overall performance of the day. Each score is a value between 0 and 6.
- Evaluation of the work
  Questions for participants to self-evaluate their work for the day. The score is based on a 5-point scale for the following questions. Each score is a value between 0 and 4 (7 questions). (1)"I was able to concentrate on my work", (2)"I was able to work efficiently", (3)"I was able to work on schedule" , (4) "I was able to communicate well with the people involved,"(5)"I was able to communicate efficiently with the people involved", (6)"I was able to come up with new ideas", (7)"I was able to achieve results"

4 Methods

In this section, we describe the analysis methodology, starting with the feature extraction and preprocessing of sensor data, behavioral data, and questionnaire data separately, followed by a modeling approach.

4.1 Preprocessing and Feature extraction

4.1.1 Sensor data

The wristband device recorded calories burned, sleep characteristics, metabolic equivalents, the number of steps, the number of floors, and activity characteristics every minute, and heart rate every 10 seconds. All of these data were aggregated for each day for each participant. We used the mean, variance, and median values for each sensor over the course of a day as features. However, since the time of day for sleep was limited, data from 0:00 to 8:00 were aggregated. In the sleep data, the sleep state was expressed in three levels (awake, deep sleep and light sleep state) so the sleep time for each sleep stage was aggregated.

4.1.2 Behavioral data

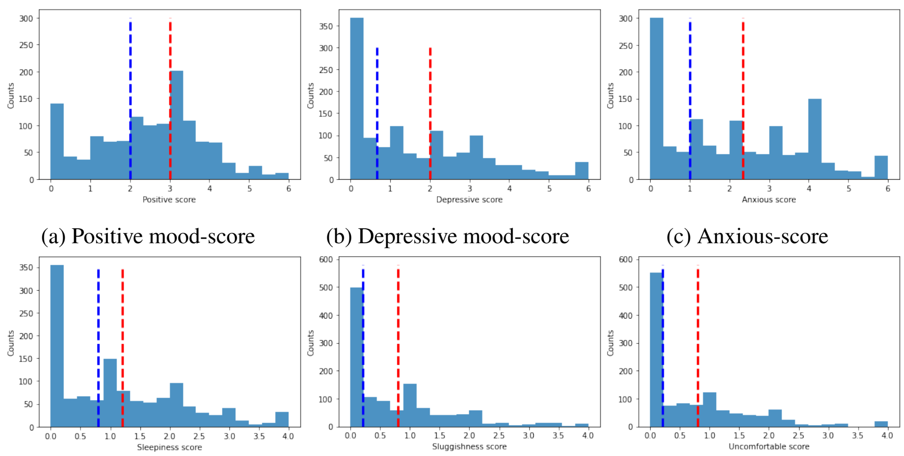
Behavioral data are the types and functions of work tasks labels described in section 3. The data were aggregated on a daily basis for each participant, and the value indicated the duration of the action in minutes. If the action was not performed, the value 0 (minutes) was entered.

4.1.3 Weather and other data

For the explanatory variables, in addition to the sensor data and behavioral data collected in the experiment, we added weather data, day of the week, and participant information for prediction models. Weather information is important not only because it directly affects people's psychology, but also because it affects when participants exercise and study [37]. Weather data were collected from the Japan Meteorological Agency's website based on the participant's residential area. The data were aggregated for each day, and the features below were computed. Average temperature, maximum temperature, minimum temperature, precipitation, sunshine duration, average wind speed, maximum wind speed, average pressure, cloud cover. The participant information included age and gender.

4.1.4 Questionnaire data

Each psychological index was scored for each questionnaire. The scoring method for each psychological indicator is shown in section3.3. The scores were computed in different ways for each psychological indicator and were continuous values. We considered the two classes of data, which were the lower 40% and the upper 40%, because we designed binary classification of each psychological trend which was either high or low. The data in the intermediate 20% bins were deleted. This labeling method was based on [37]. Figure1 shows the histogram of the psychological indices of "DAMS" and "Subjective Pain" collected from the morning questionnaire, and Figure 2 shows the histogram of the psychological indices of "DAMS", "Subjective Pain", "Recovery Experience", and "Recovery Experience" collected from the weekday evening questionnaire. The blue dotted lines on the histogram represent the 40th percentile and the red dotted line represent the 60th percentile. Therefore, the values below the blue dotted line are the lower label and the values above the red dotted line were the upper label. The histogram without the blue dotted line shows the same value for the 40th percentile and the 60th percentile. For those indicators, the mental score ≥ 60 percentile was used as the upper label and the others as the lower label.



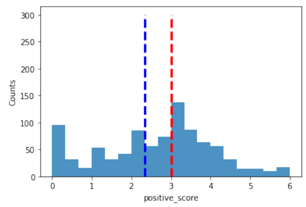
(a) Positive mood-score (b) Depressive mood-score (c) Anxious-score

(d) Sleepiness-score (e) Sluggishness-score (f) Uncounfortable-score

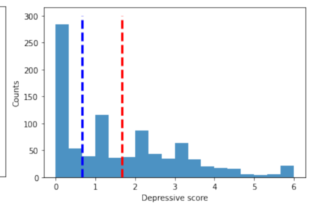
Fig. 1: Histograms show the distributions of scores on psychological indicators from the morning questionnaire. The blue dotted line indicates the 40th percentile of each psychological score, and the red dotted line indicates the 60th percentile of each psychological score. (y-axis: data counts, x-axis: score for each indicator)

4.2 Model Development

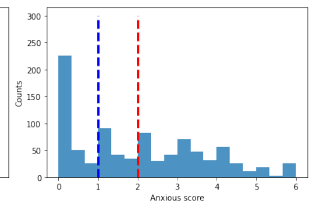
For model development, we use the same day's all explanary data when we detect the evening questionnaire and we use the previous day's explanary data when we detect the morning questionnaire. LightGBM (Light Gradient Boosting Machine) was employed as the classification model. LightGBM is a tree-based learning framework for gradient boosting. It has been created to be distributed and efficient, with a faster training speed and greater efficiency. This learning model use Gradient Boosting Decision Tree (GBDT), which is a model using Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB) in comparison to the conventional



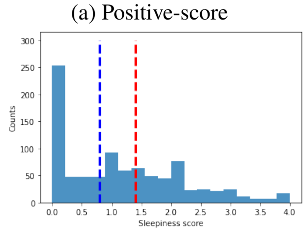
(a) Positive-score



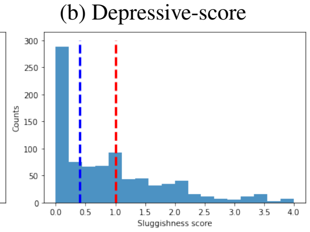
(b) Depressive-score



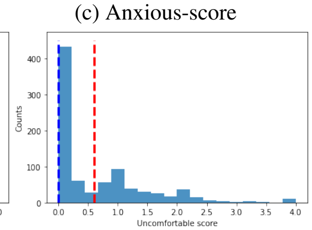
(c) Anxious-score



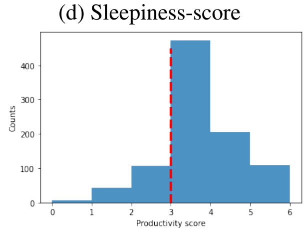
(d) Sleepiness-score



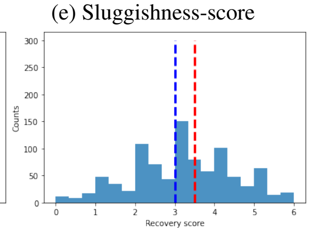
(e) Sluggishness-score



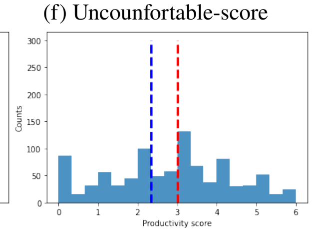
(f) Uncounfortable-score



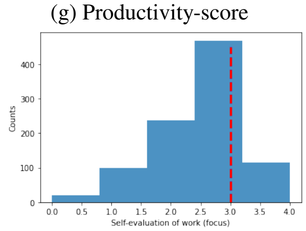
(g) Productivity-score



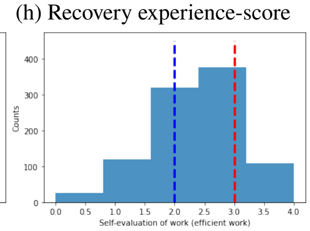
(h) Recovery experience-score



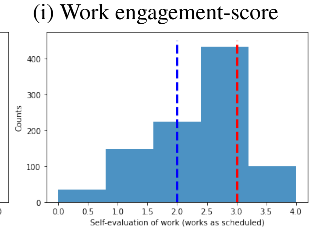
(i) Work engagement-score



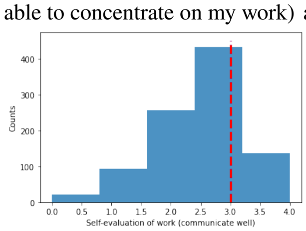
(j) Self-evaluation-score (I was able to concentrate on my work)



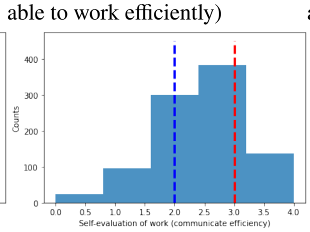
(k) Self-evaluaiton-score (I was able to work efficiently)



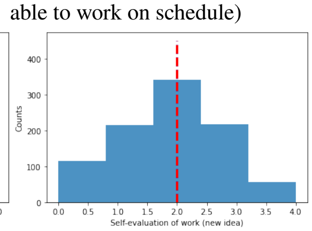
(l) Self-evaluation-score (I was able to work on schedule)



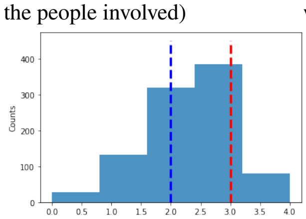
(m) Self-evaluation-score (I was able to communicate well withthe people involved)



(n) Self-evaluaiton-score (I was able to communicate efficiently with the people involved)



(o) Self-evaluation-score (I was able to come up with new idea)



(p) Self-evaluation-score (I was able to achieve results)

Fig. 2: Histograms show the distributions of scores on psychological indicators from the weekday evening questionnaire. The blue dotted line indicates the 40th percentile of each psychological score, and the red dotted line indicates the 60th percentile of each psychological score. (y-axis: data counts, x-axis: score for each indicator)

GBDT model [38]. This mechanism enables accurate estimation of small amount of data and it is a suitable algorithm for handling sparse data. Hyperparameters for each classification model were set using grid search. The model was evaluated by 5-fold cross-validation. The dataset was randomly divided into five parts, four of which were used as training data and one as validation data. We repeated the process five times and evaluated the average of the five times.

4.3 Gini Index

We can analyze how much the feature selection contributes to the classification of the target using the Gini index. First, to calculate the Gini index, we calculate the Gini impurity. The Gini impurity is expressed by the following Equation1, and is a measure of how poorly a target is classified. G(k): Impurity at a given node k, n: Number of target labels, p(i): frequency of target label i at some node k

$$G(k) = \sum_{i=1}^{n} p(i) \times (1 - p(i)) \tag{1}$$

The Gini importance is calculated based on the Gini impurity. This index shows "how much the Gini impurity can be reduced by dividing by a certain feature". The importance of a feature j is defined by Equation2 below. F(j): The set of nodes into which a feature j is to be split, N_{parent}: Samples counts at a node i, $N_{leftchild}, N_{rightchild}$: Samples counts on the left (right) side among the child nodes of a node N_{parent}: Gini impurity at a given node i, $G_{leftchild}, G_{rightchild}$: Gini impurity at the left(right)-hand side of the child nodes of a node i.

$$\begin{aligned} I(k) = \sum_{i=1}^{n \in F(j)} (N_{parent}(i) \times G_{parent}(i)) - (N_{leftchild}(i) \times G_{leftchild}(i) + \\ N_{rightchild}(i) \times G_{rightchild}(i)) \end{aligned} \tag{2}$$

In this study, we used Gini importance as the feature importance. Gini importance is a feature importance used in decision tree models, and is the weighted sum of the reduction in impurity of a node averaged over the entire decision tree. It is a measure of how much the impurity of a node is improved by using that feature.

4.4 SHAP Value Analysis and Comparison

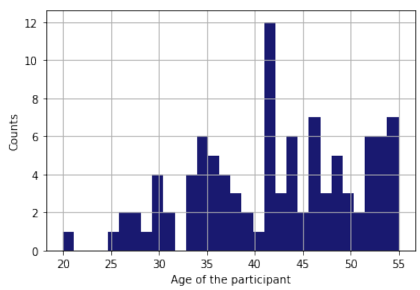
It is important to interpret how the machine learning model makes its decisions in mental health prediction because it allows us to understand how each feature affects the participant's mental health. To interpret the model, several studies have begun to use SHAP (SHapley Additive exPlanations) [39][40]. SHAP values are the contribution of each feature as determined by cooperative game theory[41]. The importance of a feature can be defined as the increase in the prediction error of the model after permuting the values of the feature. Thus, a feature can be considered important if its error increases after permuting it. If changing its value does not change the error, it is unimportant because the model is ignoring the feature in its decision. This method allows us to quantify the contribution of a feature, and its consistency in the classification prediction task has been demonstrated[42]. We analyzed the relationship between the explanatory and target variables by interpreting the model predicting the psychological indicators using their SHAP values. The SHAP values give an idea of the additivity of the features to the explanatory variables. It shows how much each feature positively or negatively affects the model. Specifically, we examined the relationship between office workers' behavior and their psychological state, as well as workplace factors that influence their mental health. Furthermore, we analyze work-related factors that affect the psychological indicators of office workers. For example, in terms of behavioral states, we analyzed if the time spent on "face to face meetings", "eating," "working alone," "hobbies," "resting,", "traveling", "web conference", "collaborative work" and other behavioral affect the different psychological state e.g., positive mood (morning and night), depressive mood (morning and night) and anxious mood (morning and night). In addition, "with whom the work was done" and "whether the work was standardized" were also analyzed with the different psychological states.

5 Results

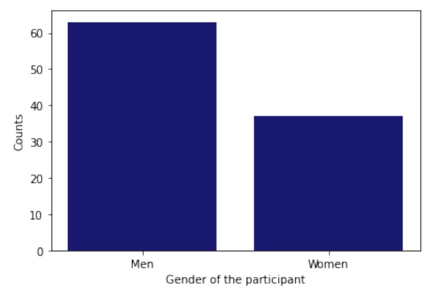
We present the results from three analyses: (1) data statistics, (2) psychological index prediction performance, and (3) contributions of participants' work tasks and environment to their psychological measures: positive mood, depressive mood and anxiety mood of the DAMS questionnaire described in section3.

5.1 Data Statistics

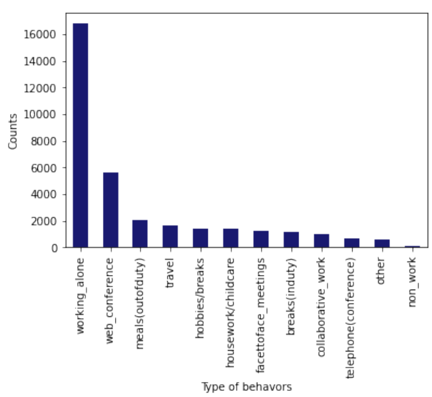
Figure 3 shows data statistics by age, gender, type of behavior, and questionnaire for the collected data. We have different work tasks (Fig 3 (c), such as working alone, working via web conferencing, collaborative work, mealtime, hobbies, and so on. Working alone was the most frequent behavior label. Figure 4, represents data attributes based on work task details and work environment. Figure 4 (a) and (b), are the representation of core or non-core work as well as standardized and non-standardized work. Figure 4 (c) shows types of tasks in the office for example, office tasks, development tasks, planning tasks, manager tasks etc. The histogram of the participants' workplaces is shown in Figure 4 (d). Because these data were collected during the Covid-19 pandemic time, more data were collected at home, not in an office. Similarly, Fig (e) depicts data counts based on the participants' roles in tasks, while Fig 4 (f) depicts data counts about whom the participants worked with. Our data show that the distributions of age and gender are not uniform. The participants were randomly selected, and we believe that the distributions match ones of workers who perform office work in a realistic manufacturer IT company. As seen in Fig 4, there was imbalance in the amount of data for each task type and environment. We believe that these are also imbalances that could be expected in the real world. For example, as shown in (a), the number of core tasks was probably much higher than the number of non-core tasks, and it is easy to imagine that many office workers work alone.In addition, since the values of each data used in this study are based on the working hours worked for each variable, we believe that the difference in the amount of data between each feature is not a problem in this analysis.



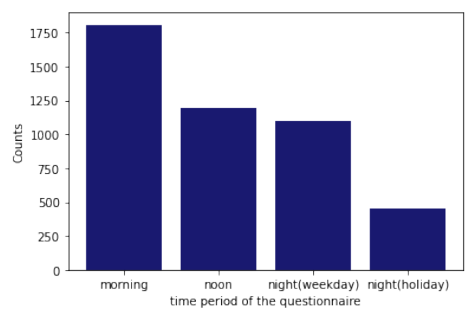
(a) Participants' age



(b) Gender of participants

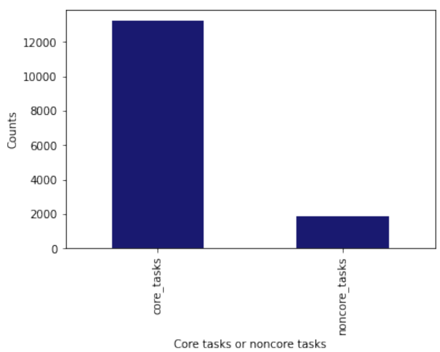


(c) Behavior Labels

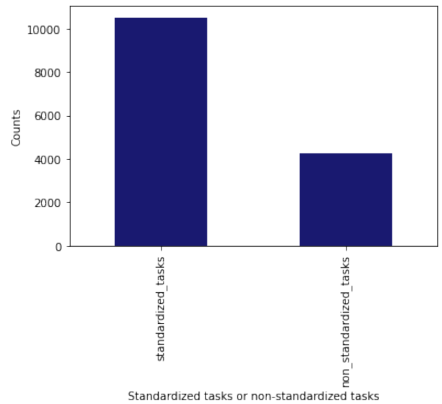


(d) Questionnaires filled out at different time period of the day

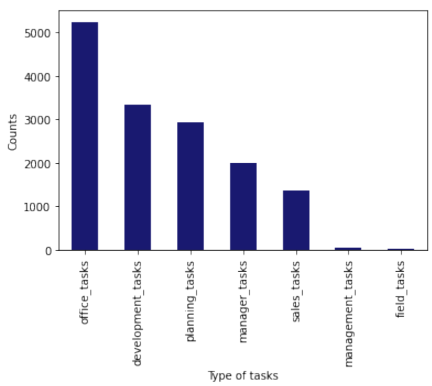
Fig. 3: Histograms of data collected in the experiment



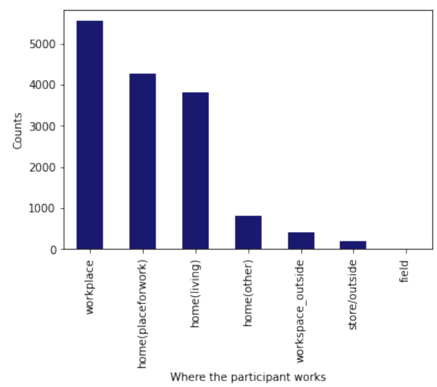
(a) Core task vs non-core tasks



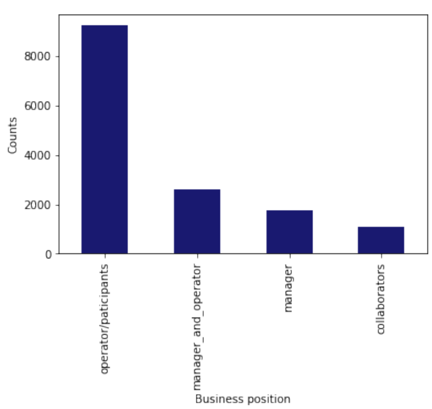
(b) Standardized vs non-standardized tasks



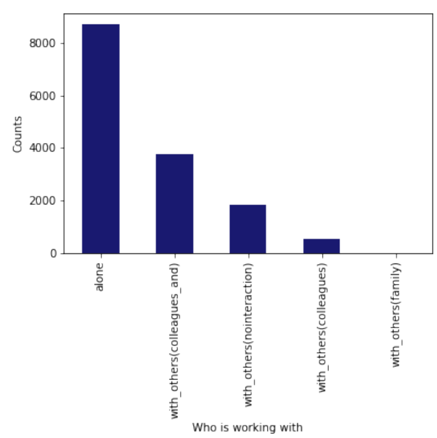
(c) Participant's task types



(d) Location of participants' work



(e) Participants' business positions



(f) Work collaborators

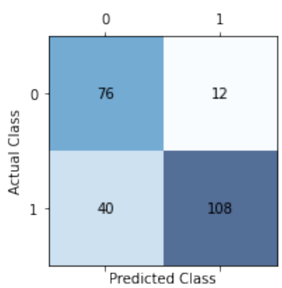
Fig. 4: Work details and environment

5.2 Analysis 1: Predictive Results of Psychological Indicators

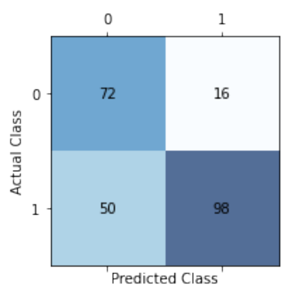
In this section, we show the results of classification prediction for each questionnaire data in binary class as shown in section 4.2. For the morning questionnaire prediction, the data collected on the previous day were used for the explanatory variables, and the data collected from the night of the previous day to the morning of the previous day were used only for the sleep features.

The performance evaluation of the classifier was performed using Accuracy and F1 Score. The prediction results are shown with all 75 feature values (Work task related features: 40, Participant-related features: 2, Weather: 9, Sensors: 24) as the target variables first. Then, features were selected and the result was shown with 22 target variables as we predicted 6 mental indicators on the morning questionnaire and 16 mental indicators on the evening questionnaire. The benefits of prediction with feature selection includes (1) it is possible to eliminate the the number of features which becomes noise and improve prediction accuracy, (2) it reduces the memory and time required for learning.

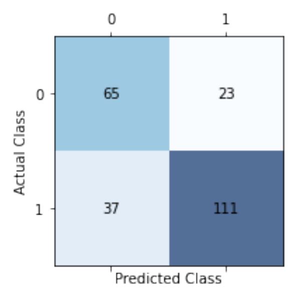
In order to see how much behavioral data, weather data, and sensor data each contribute to the prediction, we showed the prediction of positive mood index using only a single modal of data in Table 3. For example. we first used only sensor data, then work data (behavior data), and afterwards we used weather data to predict positive mood. The confusion matrix is also shown in Figure 5.



(a) Confusion matrix for predicting morning positive mood using only behavior data



(b) Confusion matrix for predicting positive morning mood using only weather data.



(c) Confusion matrix for predicting morning positive mood using only sensor data.

Fig. 5: Confusion matrix for predicting positive mood in the morning using various data.

The predicted results for the morning questionnaire are shown in Table 4, and the predicted results for the evening questionnaire are shown in Table 5, where all the data collected in this study were used as input. Comparing Table 3, Table 4, and Table 5, we can see that each type of data contributes to the prediction, and the prediction accuracy improved when multiple data were used at the same time.

Table 3: Morning questionnaire prediction results with single modal data

Mental Indicaters	Objective Variable	Type of data	accuracy	F1 score
DAMS	Positive Mood	Behavior data	77.2%	81.2%
		Weather data	71.0%	75.2%
		Sensor data	75.2%	80.0%

Table 4: Morning questionnaire prediction results with all data as the target variable on the left. Prediction accuracy with all data as a target variable, right side features selection. Prediction accuracy with 30 data as target variables.

Mental Indicaters	Objective Variable	accuracy	F1 score
DAMS	Positive Mood	82.5% / 83.9%	86.2% / 87.2%
	Depressed Mood	87.1% / 88.1%	86.2% / 87.2%
	Anxious Mood	86.0% / 89.0%	85.7% / 88.7%
Subjective Pain	Sleepiness	84.9% / 86.3%	84.9% / 86.2%
	Uncomfortable	87.1% / 88.4%	86.4% / 87.8%
	Sluggishness	84.6% / 85.3%	82.4% / 83.2%

Table 5: Evening questionnaire prediction results with all data as a target variable. On the left side is the prediction accuracy with all data as the target variable, and on the right side is the prediction accuracy with 30 data as the target variable by feature quantity selection

Mental Indicaters	Objective Variable	accuracy	F1 score
DAMS	Positive Mood	72.1% / 72.1%	83.8% / 83.8%
	Depressed Mood	85.9% / 86.8%	87.8% / 88.4%
	Anxious Mood	83.8% / 85.0%	88.4% / 89.1%
Subjective Pain	Sleepiness	76.8% / 77.2%	72.4% / 72.1%
	Uncomfortable	83.2% / 82.7%	76.1% / 74.7%
	Sluggishness	81.4% / 82.9%	75.6% / 78.0%
Productivity	Mean Score	80.3% / 80.3%	79.2% / 79.2%
Recovery Experience	Mean Score	80.4% / 79.7%	80.4% / 79.6%
Work Engagement	Mean Score	79.5% / 79.4%	84.1% / 84.6%
Self-evaluation of work	(1)	81.3% / 81.1%	75.6% / 74.5%
	(2)	82.6% / 83.3%	86.2% / 87.1%
	(3)	80.2% / 81.5%	82.5% / 83.6%
	(4)	82.3% / 81.8%	85.4% / 85.2%
	(5)	80.2% / 81.8%	82.0% / 85.2%
	(6)	80.3% / 80.2%	86.3% / 86.5%
	(7)	80.2% / 79.7%	86.5% / 81.4%

5.3 Analysis 2: Relationship between psychological indicators and behavior

In this study, we aim to clarify the relationship between each psychological index and the behavior of office workers. In this section, we predicted participants' mood by using behavioral data as a target variable and showed the relationship between data from the SHAP value and feature importance. The correlation between the work data and the psychological data may not have been correctly analyzed because age and gender acted as confounding factors. Therefore, only the work data and psychological data were used for Analysis 2. We built a machine learning model to predict morning and weekday night DAMS in order to calculate and analyze SHAP values. Since DAMS has three indices for "positive mood," "depressed mood," and "anxious mood," we made a total of six predictions.

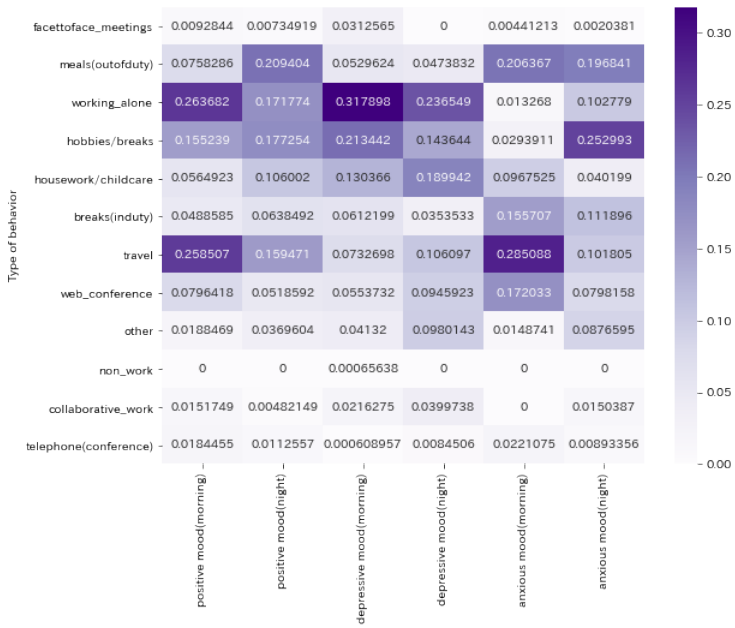
In Figure 6a and 6b, we showed which variables contributed to the prediction by visualizing the importance of the variables in predicting each psychological indicator. Figure 6a shows the variable importance in the prediction of the explanatory variables for the types of behaviors, where the y-axis shows the respective explanatory variables and the x-axis shows the objective variables. The variable importance was normalized for each objective variable, and the higher the value, the higher the importance and the darker the color of the heat map. Therefore, it is easy to see the importance of each explanatory variable in the prediction. Figure 6b shows results about feature importance analysis using behavioral details as explanatory variables, and shows the same heatmap of variable importance as in Figure 6a.

Next, in order to understand the contribution and correlation of each explanatory variable to the prediction, summary plots Were used to show SHAP Value as a one-axis scatter diagram for each feature quantity. Figure 7 shows a scatter chart of SHAP Value when the type of behavior of the participant was used as an explanatory variable. The features on the vertical axis were sorted by the mean absolute value of the SHAP values. Each plot shows the SHAP values, and the color indicates the magnitude of the feature value (blue is low, red is high). In other words, the farther the plot was from zero, the more influence it had on the inference. The features on the vertical axis were sorted by the mean absolute value of the SHAP values. By observing the color and distribution of the points, we can interpret how features affect the output. This analysis targets only continuous values of features. For example, in Figure 7 (a), we observed that 'participants/operator' was at the top of the list because it showed the highest average SHAP value. Also, we can see that there are many red plots in the negative direction far from 0, and many blue plots in the positive direction near 0. Therefore, the higher the value of 'operator/participants, the more it influences the model in the negative direction (lower psychological score). A low SHAP value was also observed for a low 'operator/participants' value that affected the model in a positive direction (higher psychological score).

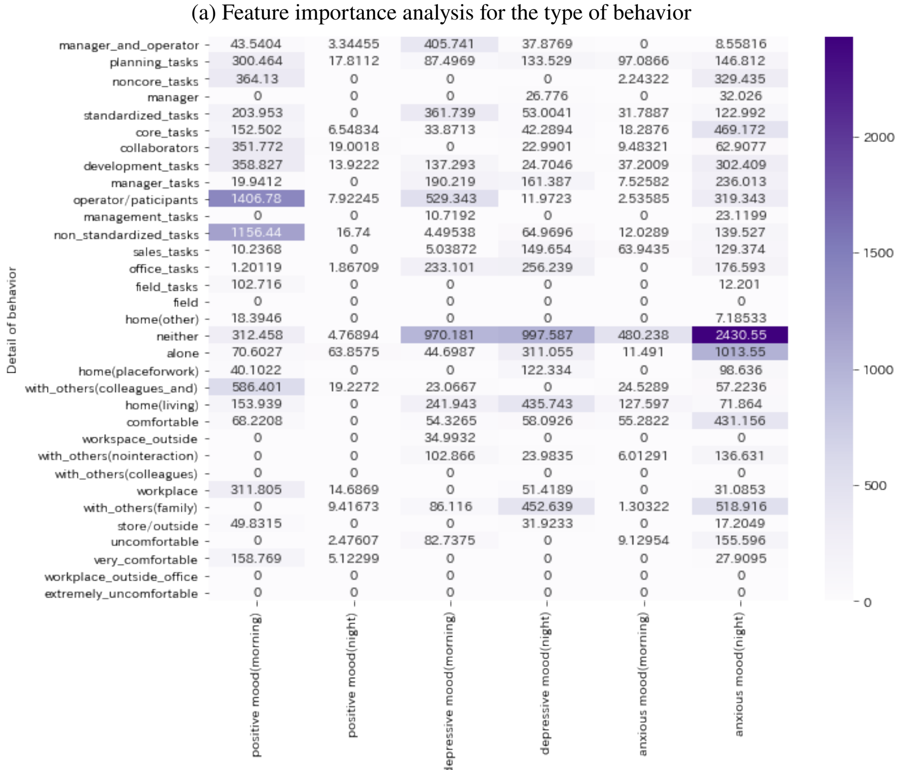
6 Discussion

6.1 Discussion for Analysis 1

The prediction accuracy scores for all psychological indicators were 72-89%, which was a stable and high accuracy overall. In our findings, the prediction results for

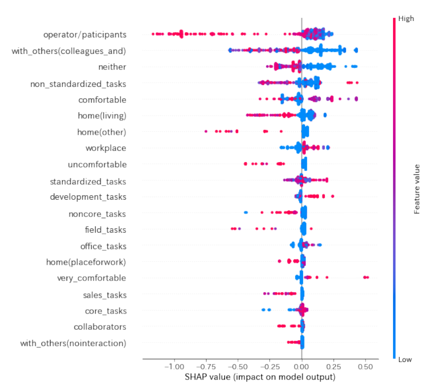


(a) Feature importance analysis for the type of behavior

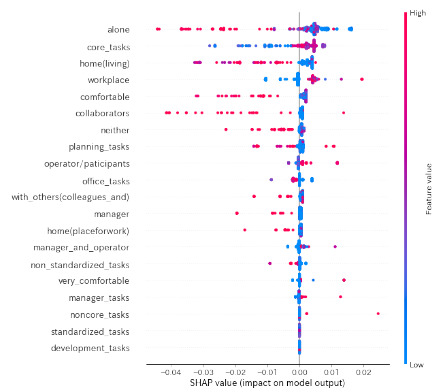


(b) Feature importance analysis with environment and details of behavior.

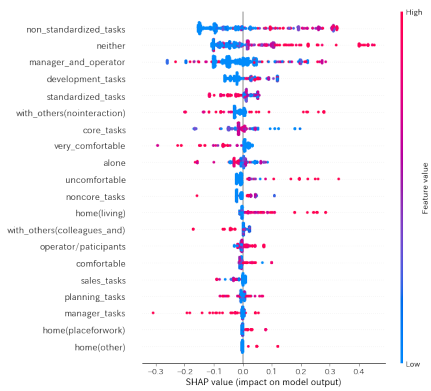
Fig. 6: The importance of each variable was calculated using the Gini importance. Each value was a normalized value for each classification model. (Y-axis: feature used for prediction, x-axis: target variable to predict)



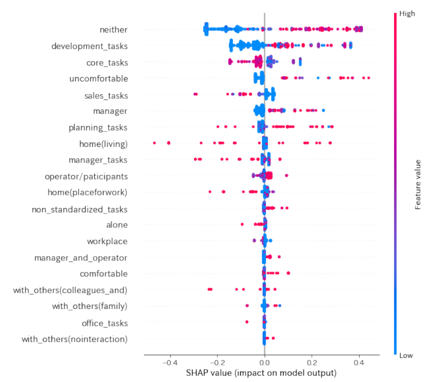
(a) positive mood prediction from the morning questionnaire



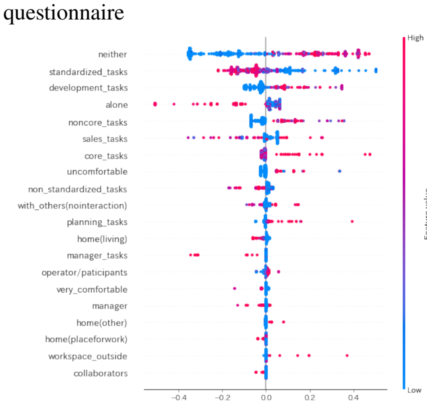
(b) positive mood prediction from the night questionnaire



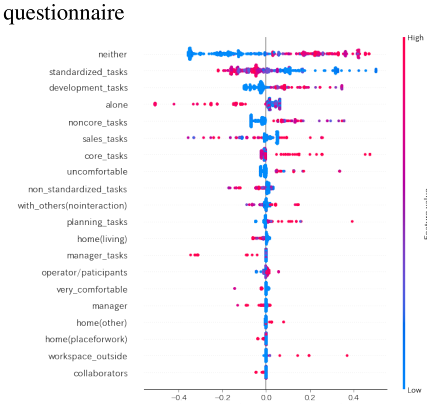
(c) depressive mood prediction from the morning questionnaire



(d) depressive mood prediction from the night questionnaire



(e) anxious mood prediction from the morning questionnaire



(f) anxious mood prediction from the night questionnaire

Fig. 7: Summary plots for predictive classification of positive mood, depressed mood, and anxious mood with work task details and environmental data as explanatory variables

subjective pain in the night survey were lower than the prediction results for the morning survey. We believe that this may be because it takes time for a person's behavior to affect subjective pain. The pain level for the morning questionnaire was predicted based on the previous day's behavior, and the one for the night questionnaire was predicted based on the behavior of the day. Therefore, we consider that subjective pain may be more strongly influenced by the behavior of the previous day than by the behavior of the day itself. In addition, the importance of each feature amount for behavior, weather, and sensor data was found to be higher and contributing to prediction for all psychological indicators. Therefore, it is expected that better prediction accuracy can be expected if the model is made with the user's personal consideration.

Each psychological score was divided into the low and high classes using top and bottom 40% of the data, but if many of the scores distributed around the 40% boundary, the balance of the classes may be skewed: for example, the ratio of low and high score classes for morning mood was positive (57:43), depressed (51:49), anxious (51: 49). This imbalance in the data may have some effect on the accuracy, but it was not expected to be significant.

6.2 Discussion for Analysis 2

The correlation between behavior and psychological state of office workers and the factors in the workplace that affected the psychological states of office workers were investigated in this paper. Figure 6a and Figure 6b depicted the importance of variables in predicting the DAMS score using the type of behavior and detailed data with the time of the behavior as explanatory variables. They were computed using the Gini coefficient for the classification model was trained. They were normalized for each prediction. In terms of behavioral features, the time spent on "meals," "working alone," "hobbies," "break," and "traveling" had a high effect on the psychological situation. For example, if we check working_alone behavior in the Figure 6a and its corresponding values for psychological states then we can see that this behavior triggered depressive mood in the morning and depressive mood at night. In addition, "with whom the work was done" and "whether the work was standardized" were found to have high effects in psychological states. In addition, the relationship between each psychological index and the explanatory variable can be found in Figure 7. In the relationship with DAMS, positive mood and depressive mood was related to features in the morning and at night. For example, positive mood was related to "independent work", "hobby/break", "movement", and depression mood was related to "independent work", "hobby/break", "housework/child-rearing". In Figure 7 (c), (d), (e), (f), the contribution of "neither" feature for the work environment comfort evaluation was high in the work content, and all of them showed positive correlation with negative mood. Participants who answered "neither" in the work environment showed a high tendency to be depressed but were accustomed to spend more time on the work. This result suggests that the amount of time spent at work may have more significant effects on the state of mood than the quality of the work environment. Negative correlations were seen with the morning depression and the morning and night anxiety. This shows that the standardized nature of work tasks has a strong influence on the anxiety and depressive psychology of employees, and the more standardized it was, the less psychological burden they had.

Regarding the work environment in telework situations, from Figure 7 (a) and (c), there was a negative correlation between "home (living)" and positive mood at night, and there was a positive correlation with morning depression. In other words, the work environment at home also showed a strong psychological impact, and workplaces such as living tended to be prone to depression.

Using the summary plots, we also analyzed the type of work behavior and found that "web conferencing" showed a negative correlation with positive mood and a positive correlation with depressive mood. In other words, if the time of web conferencing was long, there was a high possibility that it would negatively affect the psychological state of office workers.

In the previous study [20], it was shown that subjects tended to maintain a high level of concentration when the tasks they performed were standardized. Therefore, the result that workers' psychological burden tends to be less when the tasks are standardized may be interrelated with the fact that the tasks are easy for the subjects to concentrate on. In the present analysis, days with a lot of 'travel' time showed a tendency to have a lower positive mood. Previous studies have also shown that long commute times can cause stress and tension [43], which can lead to the inability to fulfill responsibilities outside of work, resulting in lower job satisfaction [44]. As far as we know, there is no past research that revealed the relationship between workers' workplace and their psychology in an environment where remote work and office work were mixed. In addition, since behaviors and work contents were collected, we were able to analyze the relationship between the work environment and situation and the psychological situation. As limitations, however, while numerous research have used sensors and behavioral data to predict psychological indicators, there have been few studies that have tested and analyzed these findings in other real world scenario. We need to evaluate whether these findings can be generalized in other populations [45].

7 Conclusion

We conducted a 14-day experiment to collect wrist-band sensor data as well as behavioral and psychological questionnaire data from about 100 office workers. By designing machine learning prediction models using behavioral data, sensor data, and weather data, we were able to classify office workers' high/low six psychological indicators with 72-89% accuracy. The main findings obtained from the experiments and discussions in the paper are shown below:

- When binary classification of psychological indicators was performed using all the data collected in this experiment, the F1 score of 17 items was 80% or more

accurate in the prediction of 22 items of psychological indicators, and the accuracy was higher compared to the previous study [14][18][37].
- Comparing the prediction accuracy of the morning questionnaire and the evening questionnaire, the accuracy of the night subjective pain examination was significantly reduced.
- Correlations between office workers' work behavior and psychological state were also investigated to reveal the factors in the workplace that affected office workers' psychological state. As a result, we found that some factors like ' 'web conferencing', working at 'home (living room)' and 'break time (work time)' had a significant impact on the psychological state. For example, 'web conferencing' has a negative correlation in a positive mood and a positive correlation in a depressive mood. The correlation between 'break time (work time)' and each mood was also observed.

In the future, we will understand individual differences among office workers. The information such as different personality characteristics might improve prediction accuracy. The prediction accuracy can be further improved by using a model that takes into account the time relationship of a day. Previous studies have also improved the accuracy of models that take into account the time series of sensor data and behavioral data [1][17]. Also in this paper, we only used the duration of the behavior as features, so we will add the time-series relationship and frequency of the work behavior in the future to get more insights into the data. In the future, other machine learning models and ensemble learning approaches should be considered for prediction improvement.

8 Acknowledgments

The experiment of this research was carried out in collaboration with companies and universities participating in the "2020 Sensing Transformation Study Group", whose secretariat is the Applied Brain Science Consortium.

References

1. Boning Li and Akane Sano. Extraction and interpretation of deep autoencoder-based temporal features from wearables for forecasting personalized mood, health, and stress. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 4(2), June 2020.
2. Vedant Das Swain, Koustuv Saha, Hemang Rajvanshy, Anusha Sirigiri, Julie M. Gregg, Suwen Lin, Gonzalo J. Martinez, Stephen M. Mattingly, Shayan Mirjafari, Raghu Mulukutla, Subigya Nepal, Kari Nies, Manikanta D. Reddy, Pablo Robles-Granda, Andrew T. Campbell, Nitesh V. Chawla, Sidney D'Mello, Anind K. Dey, Kaifeng Jiang, Qiang Liu, Gloria Mark, Edward Moskal, Aaron Striegel, Louis Tay, Gregory D. Abowd, and Munmun De Choudhury. A multisensor person-centered approach to understand the role of daily activities in job performance with organizational personas. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 3(4), December 2019.

3. Weichen Wang, Shayan Mirjafari, Gabriella Harari, Dror Ben-Zeev, Rachel Brian, Tanzeem Choudhury, Marta Hauser, John Kane, Kizito Masaba, Subigya Nepal, Akane Sano, Emily Scherer, Vincent Tseng, Rui Wang, Hongyi Wen, Jialing Wu, and Andrew Campbell. *Social Sensing: Assessing Social Functioning of Patients Living with Schizophrenia Using Mobile Phone Sensing*, page 1–15. Association for Computing Machinery, New York, NY, USA, 2020.
4. Fan Yang, Tingting Han, Kesui Deng, and Yong Han. The application of artificial intelligence in the mental diseases. In *Proceedings of the 2020 Conference on Artificial Intelligence and Healthcare*, CAIH2020, page 36–40, New York, NY, USA, 2020. Association for Computing Machinery.
5. Labor Ministry of Health and Welfare. Overview of the 2018 occupational safety and health survey (fact-finding survey). `https://www.mhlw.go.jp/toukei/list/dl/h30-46-50_gaikyo.pdf`.
6. S. Sharief. Job stress and its impact on employees' performance a study with reference to employees working in bpos m.kotteeswari research scholar,the new collage, chennai. 2014.
7. Peter Warr and Karina Nielsen. *Wellbeing and work performance*. 02 2018.
8. Maria S Kopp, Adrienne Stauder, György Purebl, Imre Janszky, and Árpád Skrabski. Work stress and mental health in a changing society. *European Journal of Public Health*, 18(3):238–244, 08 2007.
9. Peter Warr and Karina Nielsen. *Wellbeing and work performance*. 02 2018.
10. S. Sharief. Job stress and its impact on employees' performance a study with reference to employees working in bpos m.kotteeswari research scholar,the new collage, chennai. 2014.
11. Han Yu, Asami Itoh, Ryota Sakamoto, Motomu Shimaoka, and Akane Sano. *Forecasting Health and Wellbeing for Shift Workers Using Job-Role Based Deep Neural Network*, pages 89–103. 02 2021.
12. Dimitris Spathis, Sandra Servia-Rodriguez, Katayoun Farrahi, Cecilia Mascolo, and Jason Rentfrow. Sequence multi-task learning to forecast mental wellbeing from sparse self-reported data. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery Data Mining*, KDD '19, page 2886–2894, New York, NY, USA, 2019. Association for Computing Machinery.
13. Akane Sano and Rosalind W. Picard. Stress recognition using wearable sensors and mobile phones. In *Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, ACII '13, page 671–676, USA, 2013. IEEE Computer Society.
14. S. Koldijk, M. A. Neerincx, and W. Kraaij. Detecting work stress in offices by combining unobtrusive sensors. *IEEE Transactions on Affective Computing*, 9(02):227–239, apr 2018.
15. Ane Alberdia, Asier Aztiriaa, Adrian Basarabb, and Diane J. Cook. Using smart offices to predict occupational stress.
16. Akane Sano, A. Phillips, A. Z. Yu, A. W. McHill, Sara Taylor, Natasha Jaques, C. Czeisler, E. Klerman, and Rosalind W. Picard. Recognizing academic performance, sleep quality, stress level, and mental health using personality traits, wearable sensors and mobile phones. *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, pages 1–6, 2015.
17. Pablo Robles-Granda, Suwen Lin, Xian Wu, Sidney D'Mello, Gonzalo J. Martínez, Koustuv Saha, Kari Nies, Gloria Mark, Andrew T. Campbell, Munmun De Choudhury, Anind D. Dey, Julie M. Gregg, Ted Grover, Stephen M. Mattingly, Shayan Mirjafari, Edward Moskal, Aaron Striegel, and Nitesh V. Chawla. Jointly predicting job performance, personality, cognitive ability, affect, and well-being. *CoRR*, abs/2006.08364, 2020.
18. Amir M. Rahmani, Nima TaheriNejad, and Paolo Perego. A method for simplified hrqol measurementby smart devices. *Wireless Mobile Communication and Healthcare*, pages 91–98, 2017.
19. Tiantian Feng, Brandon Booth, Brooke Baldwin-Rodríguez, Felipe Osorno, and Shrikanth Narayanan. A multimodal analysis of physical activity, sleep, and work shift in nurses with wearable sensor data. *Scientific Reports*, 11, 04 2021.
20. Matthew Lee. Detecting affective flow states of knowledge workers using physiological sensors. *CoRR*, abs/2006.10635, 2020.

21. Hideyuki Fukuda, Yuri Tani, Hiroki Matsuda, Yutaka Arakawa, and Keiichi Yasumoto. An analysis of the relationship between office workers' sleep status and occupational health indicators. Technical Report 22, Nara Institute of Science and Technology, Kyushu University/JST PRESTO, nov 2019.
22. Enrique Garcia-Ceja, Venet Osmani, and Oscar Mayora. Automatic stress detection in working environments from smartphones' accelerometer data: A first step. *IEEE Journal of Biomedical and Health Informatics*, 20(4):1053–1060, Jul 2016.
23. Alexandros Zenonos, Aftab Khan, Georgios Kalogridis, Stefanos Vatsikas, Tim Lewis, and Mahesh Sooriyabandara. Healthyoffice: Mood recognition at work using smartphones and wearable sensors. pages 1–6, 03 2016.
24. Shayan Mirjafari, Kizito Masaba, Ted Grover, Weichen Wang, Pino Audia, Andrew T. Campbell, Nitesh V. Chawla, Vedant Das Swain, Munmun De Choudhury, Anind K. Dey, Sidney K. D'Mello, Ge Gao, Julie M. Gregg, Krithika Jagannath, Kaifeng Jiang, Suwen Lin, Qiang Liu, Gloria Mark, Gonzalo J. Martinez, Stephen M. Mattingly, Edward Moskal, Raghu Mulukutla, Subigya Nepal, Kari Nies, Manikanta D. Reddy, Pablo Robles-Granda, Koustuv Saha, Anusha Sirigiri, and Aaron Striegel. Differentiating higher and lower job performers in the workplace using mobile sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 3(2), June 2019.
25. Akane Sano, Sara Taylor, Andrew Mchill, Andrew Phillips, Laura Barger, Elizabeth Klerman, and Rosalind Picard. Identifying objective physiological markers and modifiable behaviors for self-reported stress and mental health status using wearable sensors and mobile phones. *Journal of Medical Internet Research*, 20, 11 2017.
26. Yoshihiko Suhara, Yinzhan Xu, and Alex 'Sandy' Pentland. Deepmood: Forecasting depressed mood based on self-reported histories via recurrent neural networks. In *Proceedings of the 26th International Conference on World Wide Web*, WWW '17, page 715–724, Republic and Canton of Geneva, CHE, 2017. International World Wide Web Conferences Steering Committee.
27. Terumi Umematsu, Akane Sano, Sara Taylor, Masanori Tsujikawa, and Rosalind W. Picard. Forecasting stress, mood, and health from daytime physiology in office workers and students. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, pages 5953–5957, 2020.
28. Yuliya Lutchyn, Paul Johns, Mary Czerwinski, Shamsi Iqbal, Gloria Mark, and Akane Sano. Stress is in the eye of the beholder. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 119–124, 2015.
29. Brad Shuck and Thomas Reio. Employee engagement and well-being. *Journal of Leadership Organizational Studies*, 21:43–58, 01 2013.
30. Yuki Matsuda Sozo Inoue Yuri Tani, Shuichi Fukuda and Yutaka Arakawa. Workersense: Mobile sensing platform for collecting physiological, mental, and environmental state of office workers. *In 2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, 2020.
31. Sozo Inoue, Paula Lago, Tahera Hossain, Tittaya Mairittha, and Nattaya Mairittha. Integrating activity recognition and nursing care records: The system, deployment, and a verification study. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 3(3), 2019.
32. Itaru Fukuda. Attempting to develop depression and anxiety mood scale(dams). *Action Therapy Research*, 23(2):83–93, 1997.
33. Tomohide Kubo, Norihide Joh, Hidemaro Takeyama, Tsuyoshi Makihara, Tatsuki Inoue, Toshimasa Takanishi, Yuko Aragomo, Motogo Murazaki, and Itani Tetsu. Examination of the expression pattern of fatigue during consecutive night shifts by.
34. Akihito Shimazu, Sabine Sonnentag, Kazumi Kubota, and Norito Kawakami. Validation of the japanese version of the recovery experience questionnaire. *Journal of occupational health*, 54, 03 2012.
35. Maria S Kopp, Adrienne Stauder, György Purebl, Imre Janszky, and Árpád Skrabski. Work stress and mental health in a changing society. *European Journal of Public Health*, 18(3):238–244, 08 2007.
36. Catherine MPA; Beck Arne PhD; Berglund Patricia M.BA; Cleary Paul D. PhD; McKenas David MD; Pronk Nico PhD; Simon Gregory MD; Stang Paul PhD; Ustun T. Bedirhan MD;

Wang Phillip MD ScD Kessler, Ronald C. PhD; Barber. The world health organization health and work performance questionnaire (hpq). *Journal of Occupational and Environmental Medicine*, 45:p 156–174, 02 2003.

37. Sara Taylor, Natasha Jaques, Ehimwenma Nosakhare, Akane Sano, and Rosalind Picard. Personalized multitask learning for predicting tomorrow's mood, stress, and health. *IEEE Transactions on Affective Computing*, 11(2):200–213, 2020.
38. Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.
39. Julio C. S. Reis, André Correia, Fabrício Murai, Adriano Veloso, and Fabrício Benevenuto. Explainable machine learning for fake news detection. In *Proceedings of the 10th ACM Conference on Web Science*, WebSci '19, page 17–26, New York, NY, USA, 2019. Association for Computing Machinery.
40. Amir Bahador Parsa, Ali Movahedi, Homa Taghipour, Sybil Derrible, and Abolfazl (Kouros) Mohammadian. Toward safer highways, application of xgboost and shap for real-time accident detection and feature analysis. *Accident Analysis Prevention*, 136:105405, 2020.
41. Scott Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *CoRR*, abs/1705.07874, 2017.
42. Scott M. Lundberg, Gabriel G. Erion, and Su-In Lee. Consistent individualized feature attribution for tree ensembles, 2019.
43. Avraham N. Kluger. Commute variability and strain. *Journal of Organizational Behavior*, 19(2):147–165, 1998.
44. Robert H. Moorman. The influence of cognitive and affective based job satisfaction measures on the relationship between satisfaction and organizational citizenship behavior. *Human Relations*, 46(6):759–776, 1993.
45. Anja Thieme, Danielle Belgrave, and Gavin Doherty. Machine learning in mental health: A systematic review of the hci literature to support the development of effective and implementable ml systems. *ACM Trans. Comput.-Hum. Interact.*, 27(5), August 2020.