Journal of Ambient Intelligence and Humanized Computing
https://doi.org/10.1007/s12652-026-05049-8

ORIGINAL RESEARCH

SenseDesk: a table-top device for activity and user recognition via center of gravity and weight analysis

Yukino Sato[1] · Kodai Ito[1] · Takumi Hasegawa[2] · Takashi Oshima[2] · Yuichi Itoh[1]

Received: 17 July 2024 / Accepted: 14 January 2026

Abstract

This paper proposes SenseDesk, a device designed to measure the center of gravity and weight on a desk to estimate the user and their desk activities. SenseDesk, equipped with four load cells on an acrylic plate, calculates the center of gravity and weight at 80 Hz and utilizes machine learning to estimate activities and users. We selected twelve desk activities, such as "Lying face down on the desk," "Using a mouse," and "Writing on paper with a pen," and evaluated their estimation accuracy. Using the random forest method, we achieved an F value of 0.775 for Within-Individual activity estimation and 0.671 for Between-Individual activity estimation. Additionally, the F value was 0.716 for user estimation. We then aggregated these activities into four groups of high-level activity categories representing common work states: "PC work," "Writing work," "Smartphone operation," and "Break." This abstraction improved the classification performance, yielding an F value of 0.888 for Within-Individual and 0.796 for Between-Individual estimation.

Keywords Work activity recognition · User estimation · Posture classification · Table-top device · Center of gravity

1 Introduction

The impact of COVID-19 has accelerated the adoption of teleworking globally (OECD 2021; Ministry of Internal Affairs and Communications 2021). Telework offers benefits such as reduced commute time and increased flexibility in terms of time and location. However, it also presents challenges like communication difficulties and the blurring of work-life boundaries (de Macêdo et al. 2020). Various Human Activity Recognition (HAR) methods have been proposed to estimate a user's state. For instance, a study utilized a camera and microphone placed in front of users to identify their posture and behavior (Jaimes 2006). However, this method can lead to excessive monitoring and psychological stress.

Murao and Terada (2016) suggested a camera-free approach using acceleration sensors attached to the user's body for state estimation. Although this method avoids visual monitoring, it may still cause physical and psychological discomfort. Alternative approaches, such as ambient sensing, have been explored to overcome these issues. Tani and Yamada (2013) employed pressure distribution on a desk to estimate activities, but using numerous sensors complicates wiring and increases installation costs.

This paper introduces *SenseDesk*, a table-top device that uses simple sensors to estimate the user and their desk activities. By using only four load cells, the system provides a stress-free and unobtrusive way to monitor the user's state. The system can also identify different activities and users.

To improve productivity in an office environment, it is important to estimate user states such as task concentration and psychological state. Monitoring desk work, estimating user states, and providing real-time feedback and

✉ Yukino Sato
yukino.sato@x-lab.team

✉ Yuichi Itoh
itoh@it.aoyama.ac.jp

Kodai Ito
kodai.ito@it.aoyama.ac.jp

Takumi Hasegawa
hasegawa.takumi@lmi.ne.jp

Takashi Oshima
oshima.takashi@lmi.ne.jp

1 College of Science and Engineering, Aoyama Gakuin University, 5-10-1, Fuchinobe, Chuo-ku, Sagamihara, Kanagawa 2525258, Japan

2 Link and Motivation Inc., Kabukiza Tower, 4-12-15, Ginza, Chuo-ku, Tokyo 1040061, Japan

Published online: 07 February 2026

environmental adjustments can improve work efficiency. Furthermore, by estimating each user's state, it is possible to provide personalized feedback, assign tasks at optimal times, and recommend breaks as needed. This leads to improvements in work efficiency, and contributes to increased productivity across the entire office.

The primary contributions of this paper are as follows:

1. We defined desk activities to be estimated through a questionnaire survey.
2. We used movements spanning several seconds as feature values instead of frame-by-frame postures, determining the optimal parameters for features such as frequency and window size.
3. We conducted both Within-Individual and Between-Individual estimations, applying models learned from one user's data to other users.

The structure of this paper is as follows. In Sect. 2, related research is discussed. Section 3 provides an overview of the implemented system and device. In Sect. 4, we conduct performance evaluation experiments on the implemented system, followed by discussions in Sect. 5. Finally, Sect. 6 concludes with a summary of this research.

2 Related work

2.1 Human activity recognition (HAR)

Several studies have explored HAR using various sensors to estimate human behavior in different contexts. HAR methods utilizing cameras (Inoue and Matsusaka 2010; Joutou and Yanai 2009) and wearable sensors (Ho and Intille 2005; Kern et al. 2007) have been proposed. However, cameras can lead to privacy concerns and psychological stress, while body-mounted sensors may limit movement and cause physical discomfort.

2.1.1 Ambient sensing

Research with ambient sensing methods has been investigated to estimate human states more naturally without causing physical or psychological stress. Some studies have used infrared sensors to detect user movements and identify occupants (Murao et al. 2012).

For example, pressure sensors installed in chairs have been used to identify posture and movement during seating (Mutlu et al. 2007). Arnrich et al. (2010) demonstrated that pressure distribution during seating can indicate stress levels. While ambient sensing can provide valuable data without user awareness, the high cost and complexity of sensor installation can be still challenging.

In recent years, research on activity recognition using deep learning and multimodal approaches has been progressing. Maradana et al. (2024) have utilized deep learning to achieve activity recognition. However, it requires a large amount of data and training time. Haresamudram et al. (2024) integrated multiple sensors and achieved activity recognition. However, there is a problem that synchronization and integration between sensors is difficult.

2.1.2 Estimation of desk activities

"iWood," an interactive plywood that can sense vibrations due to frictional charging, and proposed a method to detect various user actions and gestures from the vibration patterns (Wu and Yang 2022). It identified four gestures and twelve everyday actions and achieved an accuracy of over 90% for both actions and gestures.

However, many studies on the identification of desk activities use pressure sensors, and (Chang et al. 2006) developed a dining-type table device that can automatically record meals to improve the diet. Hernandez et al. (2014) showed that pressure changes significantly when the user is stressed and when the user is not, based on text transcription and mouse clicks using a pressure-sensitive keyboard and a capacitive mouse. Thus, by using a desk as a sensing device, it is possible to obtain information on the body without the psychological stress of being monitored or the physical burden of wearing the device. However, motion identification using pressure sensors requires a large number of sensors, which complicates the wiring and incurs high installation costs.

2.2 Sensing by center of gravity and weight

As a low-cost and robust sensing method, some studies have estimated user states using the center of gravity and weight measurements from strain gauges and load cells.

2.2.1 Activity estimation using center of gravity and weight

Zhang et al. (2020) developed a fork that detects the gestures of lifting food and the weight of each bite to help users understand their eating speed and increase their eating awareness. An inertial measurement unit was used to detect gestures and a load cell was used to estimate the amount of food. Although this method provided useful insights, the heavy fork (approximately 780 g) made normal eating difficult, similar to wearable sensors.

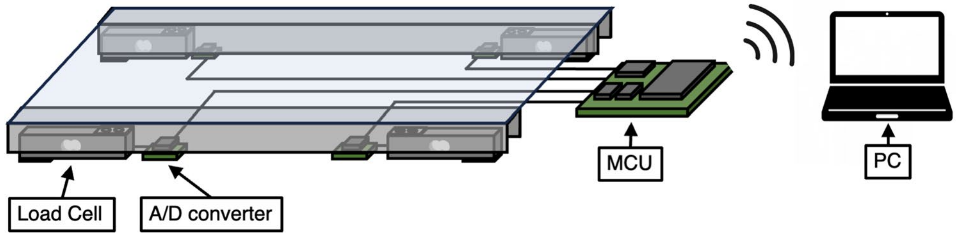

Fig. 1 System configuration

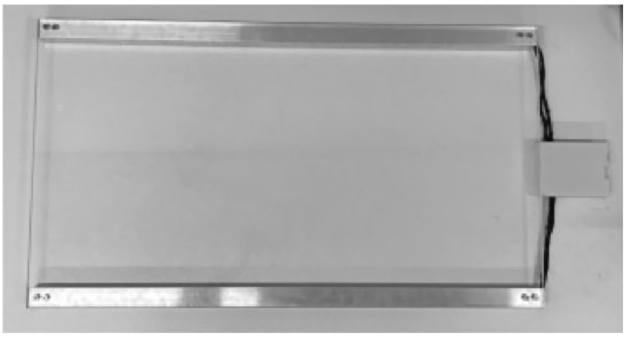

Fig. 2 SenseDesk

Kitabayashi et al. (2011) used body sway measurements by a center of gravity sway testing system to assess changes in physical condition and state. Yoshida et al. (2022) developed "Flexel," a floor interface using load cells placed at the four corners of the floor-type device to track user movements and detect object positions. Matsumoto et al. (2015) used a chair equipped with strain gauges to identify seated user states with a high accuracy of 87%.

2.3 Identification of desk activity

Schmidt et al. (2003) proposed a context acquisition system using load cells at the four corners of tables, floors, and furniture to detect object addition/removal and user movements. They also conducted research to detect events, such as touch, clicks, and tracking (Schmidt et al. 2002). Murao et al. (2017) used load cells mounted on four corners of a desk to provide three functions: object detection, action recognition, and user identification. The system was also able to recognize four types of actions with 94% accuracy and identify users with 89% accuracy for four participants and 65% accuracy for ten participants. The objective of their system is similar to that of our system. However, they identified only four activities, the criteria for which remained unclear. Moreover, their estimation was based on frame-by-frame postures for identification, neglecting user movements and behaviors. Consequently, its applicability was limited. Furthermore, because they conducted only Within-Individual estimations, their system versatility remained ambiguous. In this study, we adopted twelve desk activities through a questionnaire survey to estimate activities during desk work. For these estimations, by utilizing the features of user movements spanning several seconds, we aimed to capture the characteristics of desk activities more accurately. We also determined the optimal parameters for this purpose, such as frequency and window size. Finally, we verify the versatility of the proposed system through Between-Individual validation.

3 SenseDesk

We developed *SenseDesk*, a device that measures the series of the center of gravity and weight on a desk to estimate desk activities and identify users unconsciously. Strain gauges have excellent load-bearing capabilities and can be used to identify human conditions by using a small number of sensors. Furthermore, unlike camera-based monitoring and body-worn sensors, SenseDesk enables stress-free identification. In this section, we describe the details of SenseDesk implementation and an analysis to identify who is doing what.

3.1 Configuration of devices

Fig. 1 illustrates the system configuration, and Fig. 2 shows the developed device. The device is 10 mm thick, 800 mm wide, and 450 mm deep, which is sufficiently large to hold a keyboard and mouse. Four load cells (rated capacity of 5 kg) are placed at the four corners of the acrylic plate that served as the top panel, and all loads on the top panel are applied to these four load cells. U-shaped aluminum channels are attached to the two long sides of the top panel to reduce panel deflection. The load cell used in this study can be adapted to another desk as long as it is a level desk large enough to hold a keyboard and mouse. Furthermore, if the rigidity is high, we believe that materials such as glass, metal, and wood can be used instead of the acrylic plates used in this study.

As shown in Fig. 1, each load cell is shortly connected to an A/D converter (HX711) to prevent noise, converting weight data into digital data. An ESP32-DevKitC microcomputer with a Wi-Fi module sends the data to a PC at 80 Hz. Our system's sampling rate of 80 Hz would allow for more accurate data acquisition than Murao et al.'s system (15 Hz.)

3.2 Measurement of center of gravity and weight

The center of gravity and weight for identifying desk activities are calculated based on sensor data from the desk corners. As shown in Fig. 3, the values of the sensors at the four corners are *TopLeft*(*TL*), *TopRight*(*TR*), *BottomLeft*(*BL*), and *BottomRight*(*BR*), and the coordinates of the device center are (0, 0). If the length and width of the device are *Length* and *Width*, respectively, the center of gravity coordinates X and Y and weight W can be calculated using the following equations:

$$X = \frac{(TR + BR) - (TL + BL)}{(TL + TR + BL + BR)} \times \frac{Length}{2} \tag{1}$$

$$Y = \frac{(TL + TR) - (BL + BR)}{(TL + TR + BL + BR)} \times \frac{Width}{2} \tag{2}$$

$$W = TL + TR + BL + BR \tag{3}$$

Due to the characteristics of the sensors, an error of 5 g or 0.1% of the rated capacity may occur even when there is nothing on the desk, resulting in a total error of 20 g when four sensors are used. To avoid affecting the center of gravity measurement, the position of the center of gravity was corrected to (0, 0) if it is smaller than 20 g.

3.3 Methods of activity estimation and user estimation

To estimate desk activities and identify users using the measured data, firstly, the center of gravity X and Y and the weight W are calculated using Eqs. (1), (2), and (3) with the obtained data. Second, the data are standardized and subjected to the short-time Fourier transform while varying parameters such as the sampling rate and window size. The features include frequency components in the $X - axes$, $Y - axes$, and W, and machine learning is used for motion estimation and user identification.

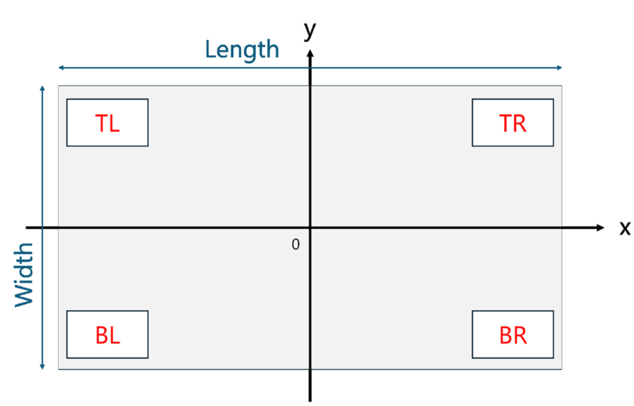

Fig. 3 Coordinates of center of gravity

4 Evaluation experiment

We conducted performance evaluation experiments to verify the accuracy of our developed table-top device in recognizing desktop activities.

4.1 Experimental methods

Twelve participants (six males and six females, age 22 ± 0.707 years, weight 57.1 ± 10.5 kg), who were undergraduate and graduate students, participated in the study. Each participant was asked to perform five sets of 10-second measurements for each activity. The data collected in the experiment included the process of beginning the movements and not only after the movements stabilized. Before each trial, the weight was reset to 0 g. However, as described above, due to the characteristics of load cells, an error of approximately 20 g can occur when no load is applied. Therefore, we collected data for 10 s after the total weight exceeded 20 g.

4.2 Target desk activities

Although Murao et al. chose only four desk activities, their selection criteria were not clarified. In this study, to ensure that our system has sufficient functions targeting desk work, we defined desk activities through a questionnaire survey. The survey involved twelve university students who reported their most frequent desk activities. Since our proposed method identifies desk activities based on their center of gravity and weight, activities not involving direct contact with the desk were excluded. Consequently, we selected the twelve desk activities listed in Table 1. In "e. Typing on a keyboard" and "f. Reading a book," participants were asked to copy Kenji Miyazawa's "Pollano Square" (Miyazawa 2005). Fig. 4 shows a scene of each activity.

Because this device identifies desk activities based on the center of gravity and weight series, participants were asked to place their hands or arms on the desk before starting each task. They were allowed to place their hands or arms in any manner and were instructed to perform their usual movements as naturally as possible.

Table 1 List for experimental task

Activity	
a	Lying face down on the desk
b	Resting cheek on left hand
c	Resting cheek on right hand
d	Using a mouse (Create shapes in PowerPoint)
e	Typing on a keyboard (Transcribe "Pollano Square" on the keyboard)
f	Reading a book
g	Writing on paper with a pen (Transcribe "Pollano Square" on paper)
h	Erasing text on paper
i	Eating a meal
j	Watching a video on a smartphone
k	Typing text on a smartphone
l	Browsing a website on a smartphone

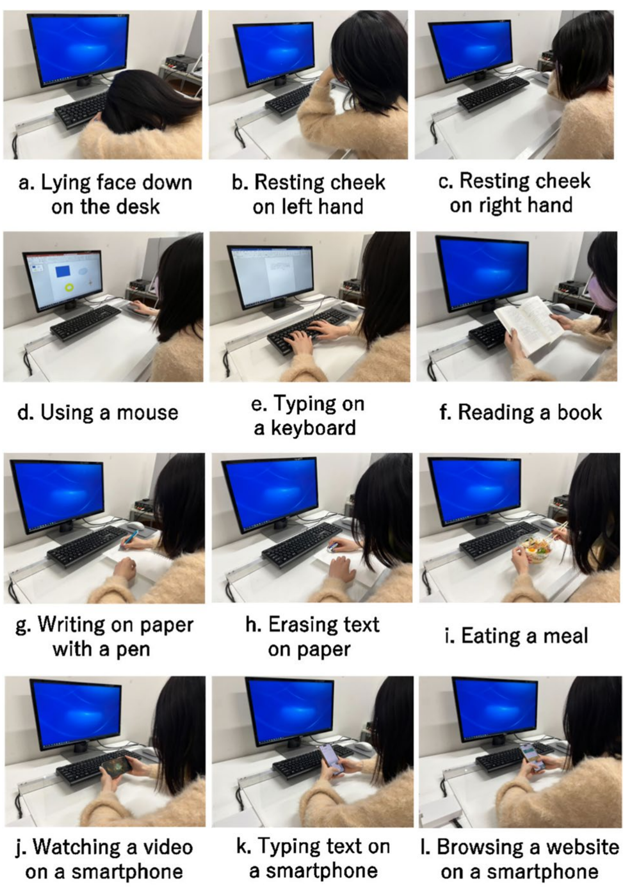

Fig. 4 Scene of each activities

4.3 Results of experiments

4.3.1 Feature extraction

We selected nine features extracted from the data: the mean and variance values for the center of gravity $X - axes$ and $Y - axes$, and weight W within a specified window, as well as values obtained by applying the short-time Fourier transform to each value. The short-time Fourier transforms were applied after resampling the data at equal intervals to account for time deviations during data transmission. The resampled data were standardized for each participant to mitigate individual differences. The short-time Fourier transform used a window size of S, sliding the window by $S/2$. The bandwidth of frequencies output by the short-time Fourier transform was divided into feature values. In addition to the output results of the short-time Fourier transform, the mean value and variance of the standardized data for X, Y, and W were calculated by sliding the window by $S/2$ with a window size S. Nine values of the mean and variance of each of X, Y, and W, and the power of each frequency component obtained by the short-time Fourier transform were calculated at sampling rates of 80 Hz, 40 Hz, and 20 Hz, and defined as features. These features were calculated for each window and used as samples for machine learning.

The parameters were compared for three different sampling rates Fs and three different window sizes S of the short-time Fourier transform; that is, nine different combinations were considered.

Table 2 Classification accuracy results for within-individual estimation

	Accuracy	Precision	Recall	F-measure
Random Forest	0.769	0.781	0.769	0.775
Decision tree	0.633	0.634	0.633	0.633
Logistic regression	0.599	0.581	0.599	0.590
k Neighbor (k = 3)	0.506	0.498	0.506	0.502
Linear SVM	0.575	0.571	0.575	0.573

4.3.2 Within-individual estimation of activity

For Within-Individual estimation, a 5-fold cross-validation was conducted with four sets of training data for each participant and one set of test data for the remaining participants, and the identification rate was calculated as the average of all participants in the experiment. Table 2 shows the highest classification accuracy of the activities among the nine parameters based on Within-Individual estimation. An F value of 0.775 was obtained using *RandomForest*. Fig. 5 shows a confusion matrix, where activities are labeled a-l as shown in Fig. 4. "d. Using a mouse" and "a. Lying face down on the desk" achieved accuracies of 99.3% and 97.6%, respectively. However, "f. Reading a book" and the three activities involving a smartphone tend to be misclassified.

4.3.3 Between-individual estimation of activity

For Between-Individual estimation, eleven participants' data were used as training data and one participant's data as test data. The process was repeated twelve times, using each

Fig. 5 Confusion matrix of within-individual estimation

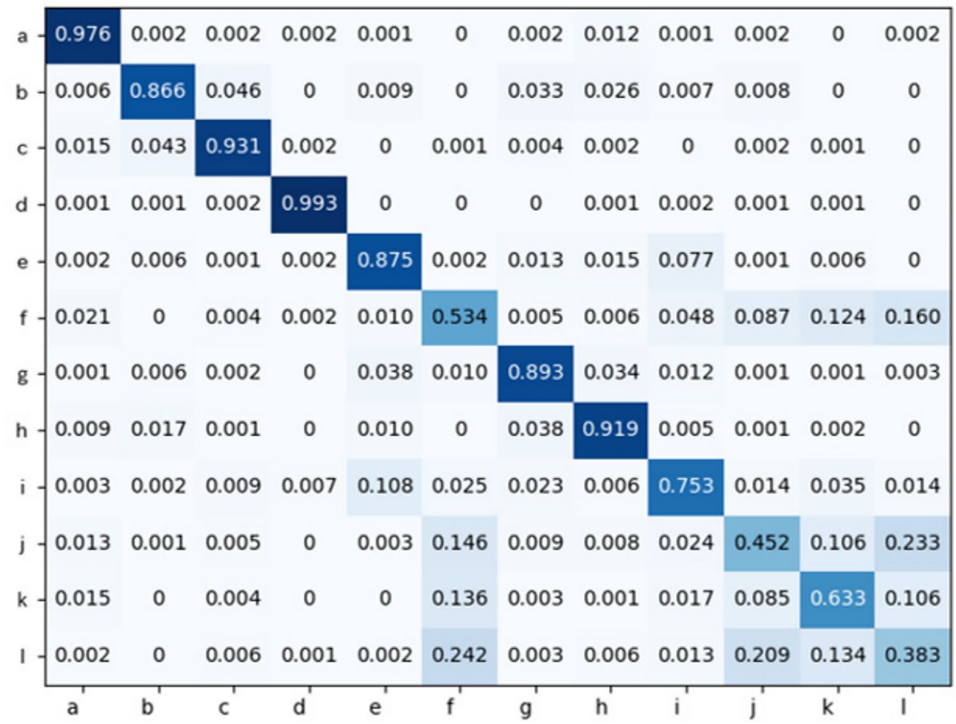

Table 3 Classification accuracy results for between-individual estimation

	Accuracy	Precision	Recall	F-measure
Random Forest	0.672	0.670	0.672	0.671
Decision tree	0.558	0.563	0.558	0.560
Logistic regression	0.537	0.528	0.537	0.532
k Neighbor (k = 3)	0.365	0.363	0.365	0.364
Linear SVM	0.516	0.516	0.516	0.516

participant as test data once, and the average identification rate was calculated. In Table 3, an F value of 0.671 was obtained by using *RandomForest*. Fig. 6 shows a confusion matrix. Classification accuracy of “c. Resting cheek on right hand” and “a. Lying face down on the desk” was 92.4% and 88.7%, respectively. On the other hand, there were many cases of misclassification within “k. Typing text on a smartphone” and “l. Browsing a website on a smartphone.” Additionally, “i. Eating a meal” was misclassified as “e. Typing on a keyboard.” In addition, compared to the fast Fourier transform and wavelet transform, the short-time Fourier transform was the most accurate for both within-individual and between-individual estimation.

4.3.4 User estimation

For user estimation, 5-fold cross-validation was performed with four of the five sets of data containing all activities of the twelve participants as training data and the remaining set as test data. In Table 4, an F value of 0.721 was obtained by using *RandomForest*. Fig. 7 shows a confusion matrix. Participants H and L were identified with high accuracies of 84.5% and 80.2%, respectively, while participants G and K had lower accuracies of 51.7% and 62.9%, respectively.

4.3.5 Parameter comparison

As mentioned above, we varied the sampling rate (80, 40, and 20 Hz) and window size (4, 2, and 1 s) for the short-time Fourier transform, resulting in nine parameter combinations.

Table 5 presents the accuracy results. The accuracy did not significantly change with downsampling. The accuracy was higher for window sizes of 2 s and 1 s compared to 4 s.

4.4 High-level activity classification through class aggregation

The accuracy mentioned above leaves room for improvement for practical applications. To address this, we performed class aggregation on the twelve types of desk activities, combining them into four high-level activity categories: “PC work,” “Writing work,” “Smartphone operation,” and “Break,” and conducted classification at this higher contextual level. The activities assigned to each category are shown in Table 6.

We performed classification for the four high-level activity categories in both Within-Individual and Between-Individual scenarios, following the same procedure used for the twelve activities. The classification was based on the parameters that achieved optimal performance in the twelve

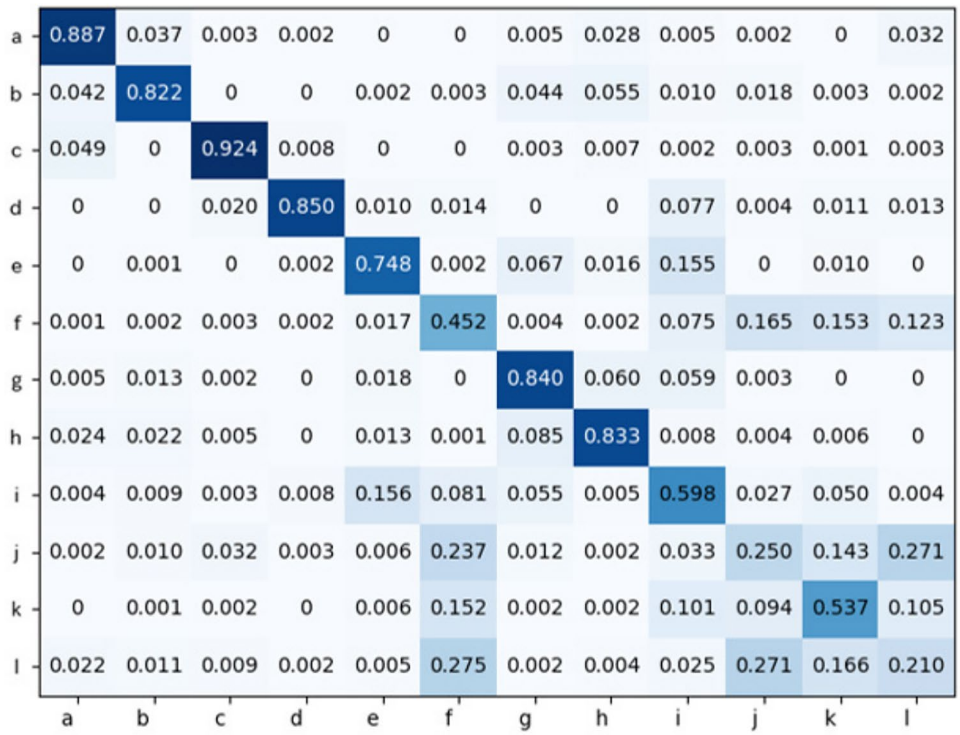

Fig. 6 Confusion matrix of between-individual estimation

Table 4 Classification accuracy results for user estimation

	Accuracy	Precision	Recall	F-measure
Random Forest	0.720	0.722	0.720	0.721
Decision tree	0.518	0.524	0.518	0.521
Logistic regression	0.364	0.355	0.364	0.360
k Neighbor (k = 3)	0.355	0.367	0.355	0.361
Linear SVM	0.501	0.493	0.501	0.497

activity recognition task: a sampling frequency of 20 Hz and a window size of 1 s. The classification accuracy for Within-Individual recognition is shown in Table 7, while Table 8 presents the results for Between-Individual recognition. Additionally, the confusion matrix for Within-Individual recognition is depicted in Fig. 8, and the confusion matrix for Between-Individual recognition is shown in Fig. 9.

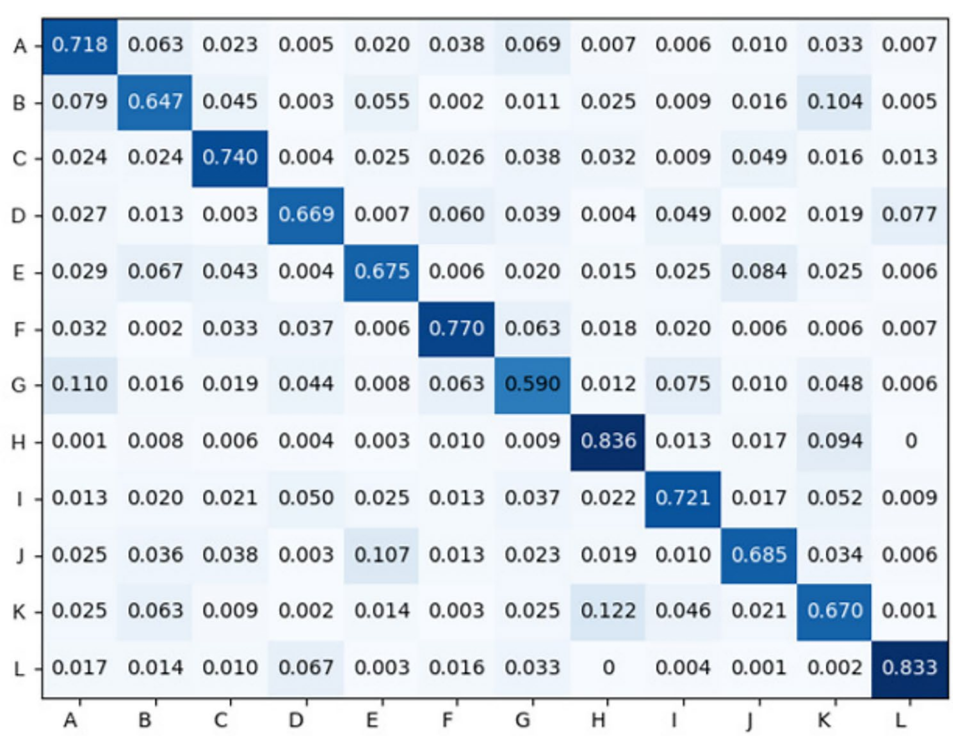

Fig. 7 Confusion matrix of user estimation

Table 5 Accuracy for each parameter

Sampling rate [Hz]	Window size [s]	Within (F value)	Between (F value)	User (F value)
80	4	0.723	0.651	0.689
80	2	0.738	0.663	0.707
80	1	0.745	0.664	0.719
40	4	0.708	0.652	0.704
40	2	0.756	0.669	0.718
40	1	0.771	0.668	0.721
20	4	0.740	0.652	0.688
20	2	0.751	0.671	0.703
20	1	0.775	0.662	0.716

Table 6 high-level activity categories

State	Activities
1. PC work	d. Using a mouse, e. Typing on a keyboard
2. Writing work	g. Writing on paper with a pen, h. Erasing text on paper
3. Smartphone operation	j. Watching a video on a smartphone, k. Typing text on a smartphone
	l. Browsing a website on a smartphone
4. Break	a. Lying face down on the desk, b. Resting cheek on left hand
	c. Resting cheek on right hand, f. Reading a book, i. Eating a meal

Table 7 Classification accuracy results for within-individual estimation in high-level activity categories

	Accuracy	Precision	Recall	F-measure
Random Forest	0.882	0.892	0.884	0.888
Decision tree	0.79	0.804	0.805	0.805
Logistic regression	0.765	0.787	0.754	0.770
k Neighbor (k = 3)	0.633	0.632	0.649	0.640
Linear SVM	0.7	0.722	0.704	0.713

Table 8 Classification accuracy results for between-individual estimation in high-level activity categories

	Accuracy	Precision	Recall	F-measure
Random Forest	0.790	0.802	0.791	0.796
Decision tree	0.704	0.709	0.706	0.707
Logistic regression	0.713	0.711	0.703	0.707
k Neighbor (k = 3)	0.578	0.580	0.583	0.581
Linear SVM	0.707	0.701	0.700	0.700

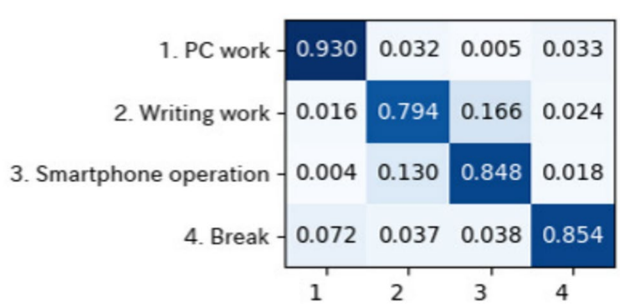

Fig. 8 Confusion matrix of within-individual estimation in high-level activity categories

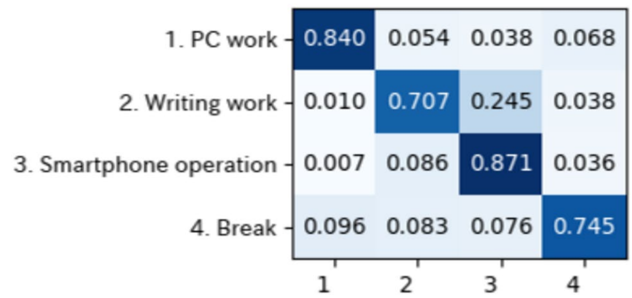

Fig. 9 Confusion matrix of between-individual estimation in high-level activity categories

The performance evaluation results indicated that the *RandomForest* classifier achieved the highest discrimination rates for both Within-Individual and Between-Individual scenarios, with F-scores of 0.888 and 0.796, respectively, consistent with the results obtained for the twelve activity classification.

In the Within-Individual scenario, the most frequent misclassifications occurred between "Writing work" and "Smartphone operation." Specifically, "Writing work" was often misclassified as "Smartphone operation." Additionally, "Break" was frequently misclassified as one of the other three high-level activity categories.

5 Discussion

5.1 Discussion of the proposed method

This study proposes a method to identify twelve types of desk activities and estimate users based on the center of gravity and weight data using the developed SenseDesk. We conducted experiments to verify the accuracy of desk activity identification and user estimation.

5.1.1 Classifier

First, we discuss the classifier results. The highest accuracy was achieved with Random Forest for both Within-Individual and Between-Individual estimations. This can be attributed to Random Forest being an ensemble learning method using decision trees, which is less susceptible to outliers. Additionally, the large variance in center of gravity and weight data across the twelve activities contributed to the method's effectiveness.

5.1.2 Within-individual estimation

For Within-Individual estimation, Random Forest achieved the highest accuracy of 0.775, with "d. Using a mouse" identified at 99.3% accuracy and "a. Lying face down on the desk" at 97.6%. "d. Using a mouse" had high accuracy due to similar activities performed within a certain range, while "a. Lying face down on the desk" was identified because of the small center of gravity shift and heavier weight. However, "f. Reading a book" and the three activities using a smartphone were prone to misclassification owing to similar arm positions and relatively static movements. Although our system achieved an estimation accuracy comparable to that of Murao et al., they estimated only four movements, whereas we achieved a high accuracy in estimating twelve movements.

5.1.3 Between-individual estimation

For Between-Individual estimation, Random Forest achieved the highest accuracy of 0.671, with "c. Resting cheek on right hand" and "a. Lying face down on the desk" were classified at 92.4% and 88.7% accuracy, respectively. The high accuracy of the former can be attributed to its unique static nature with the weight applied to one side. "g. Writing on paper with a pen" and "h. Erasing text on paper" also had high accuracies exceeding 80% owing to significant center of gravity shifts and weight changes. However, "i. Eating a meal" was misclassified as "e. Typing on a keyboard" possibly because the participants rested their arms on the desk, resembling a keyboard typing posture.

5.1.4 User estimation

For User estimation, Random Forest achieved the highest accuracy of 0.721. Participants H and L had high accuracies of 84.5% and 80.2%, attributed to distinct behaviors, such as central positioning on the desk and heavier weight on the desk. However, participants K and G exhibited lower accuracies at 62.9% and 51.7%, respectively. They were misclassified by the other participants because their movements were generally not very characteristic. These results indicate that the proposed method can achieve high accuracy in user estimation for participants with strong characteristics and habits. While Murao et al. achieved an estimation for ten users, we conducted an estimation for twelve users and achieved a higher level of accuracy.

5.1.5 Parameter comparison

In the analysis, we changed the sampling rate and the window size when performing the short-time Fourier transform. We then compared the results of the nine patterns of parameter combinations to determine the best parameter. The results showed that downsampling from 80 Hz to 40 Hz or 20 Hz did not significantly change accuracy. Window sizes of 2 s and 1 s were more accurate than those of 4 s. The frequency of the human motion is typically less than 10 Hz. Therefore, it is believed that the accuracy did not change significantly because the power of the frequencies up to 10 Hz was obtained at a sampling rate of 20 Hz. Regarding the window size, the movement time was 10 s per task, and the larger the window size, the fewer the number of samples. It is believed that higher accuracy can be obtained by increasing the number of samples per task, even if the window size is changed.

5.2 High-level activity classification through class aggregation

For Within-Individual estimation, the highest accuracy with an F value of 0.888 was obtained in the *RandomForest*. Misclassification was most frequent for "Writing works" and "Smartphone operation. This may be due to the fact that many misclassifications were made for "f. Reading a book" among the "Writing works" and for the three tasks using a smartphone, as was the case in the identification of twelve types of tasks. For Within-Individual estimation, the highest accuracy was obtained with an F value of 0.796 for *RandomForest*. Similar to the Within-Individual estimation, "Writing works" was misclassified as "Smartphone operation. In addition, "Break" was misclassified as one of the other three states. The "Break" task included the "i. Eating a meal" task, which was misclassified into various states because there were large individual differences in what people ate and how they ate it.

5.3 Privacy issues

Existing methods using cameras have privacy issues, such as the reflection of faces (Jaimes 2006). The body movement data used in this study are abstract, and if someone else is sitting at a desk, it is not possible to identify who that person is. Therefore, it can be said that privacy is currently protected compared to existing methods. However, as the accuracy of identification improves, the possibility of recognizing individuals may increase, raising new privacy concerns. As such, future work should consider strategies for anonymization, informed consent procedures, and compliance with applicable privacy regulations and ethical guidelines.

5.4 Limitation

The evaluation experiment revealed that misclassification occurred for movements with similar arm placement, and user estimation accuracy was lower for participants with fewer movement characteristics. To address these issues, the following two solutions were identified:

1. Collecting more data: In this experiment, twelve participants were asked to perform five sets of 10 s each. Therefore, increasing the amount of data for each participant and collecting data from more participants is expected to improve the discrimination rate for similar activities. In addition, there was a bias in the age and weight of the participants. Therefore, as future work, it is required to collect and evaluate data from a wider range of age groups and body sizes. Furthermore, collecting long-term data on a daily basis may better capture movement characteristics and behavior patterns.
2. Exploring more parameters: Comparing more parameters and other features during feature extraction could capture features not identified by the current method. In activity recognition and user estimation, varying the sampling rate, window size, and frequency band, as well as adding additional features, could enhance accuracy. In addition, other features not used in this study will also be considered. Furthermore, incorporating advanced feature extraction techniques and exploring sophisticated machine learning methods, such as deep learning models, are planned as part of future work to enhance classification performance.

In a natural work environment, it is also expected that people will work while switching between multiple activities, such as "d. Using a mouse" and "e. Typing on a keyboard." In such a case, the twelve categories may be too detailed, and it may be better to capture a person's state in broad categories such as "Fatigue state," "PC work," and "Writing work." These categories should be defined depending on the type of status to be estimated. In this experiment, only the center of gravity and weight were used for identification, and an F value of 0.775 was obtained for Within-individual estimation, and 0.671 for Between-individual estimation. To improve the accuracy, we will also consider using other sensors such as keyloggers, cursor movements, and smartphone sensors, as Chang et al. did with a combination of weight sensors and RFID to record meals (Chang et al. 2006). In addition, a compound operation such as "e. Typing on a keyboard" with the left hand and "d. Using a mouse" with the right hand could be expected. As a solution, it may be possible to estimate the feature values of complex motions by synthesizing the multiple feature values of a single motion.

5.5 Application

This study achieved desk activity estimation, and capturing more detailed features could enable the estimation of the user's mental state without wearable devices or cameras. This system contributes to the creation of a more adaptive and health-conscious work environment by identifying suboptimal postures and evaluating fatigue and stress levels. Previous studies have shown that a person's mental state may be reflected in their movements (Arnrich et al. 2010; Hernandez et al. 2014; Kitabayashi et al. 2011). Although this experiment was conducted in a defined environment, we plan to conduct long-term evaluations in natural environments such as actual office environments in the future. Installing this system in an office could visualize employees' conditions and fatigue levels.

One application is to visualize the time spent. By recording their own work as data, employees can objectively see how much work they do and what kind of work they have done, potentially improving work performance.

The second possible application is the estimation of the degree of concentration at work. By continuously collecting data over long periods, it may be possible to assess work progress through self-assessment labeling. If work is not progressing well, breaks can be encouraged to maintain good working conditions.

Third, when users behave unusually, their psychological state can be checked, and countermeasures such as breaks can be proposed. Furthermore, long-term data collection may predict decreases in psychological safety and suggest preventive measures, improving mental and physical health, concentration, and work efficiency.

Furthermore, because the center of gravity and weight are affected by the user's weight, sitting posture, and other factors, the generality of the model is an issue. However, we believe that when this system is introduced to offices in the future, it will be possible to achieve highly accurate classification by creating a model for each individual user rather than a single model for all users.

6 Conclusion

This study proposes a method to measure the center of gravity and weight on a desk using load cells to estimate the user's state unconsciously. We developed a table-top device, *SenseDesk*, and used machine learning to identify twelve desk activities and users. Using the Random Forest method, we achieved an F value of 0.775 for Within-Individual activity estimation, 0.671 for Between-Individual activity estimation, and 0.716 for User estimation. The Classification accuracy of the twelve activities was not very high. However, an F value of 0.888 for Within-individual estimation and 0.796 for Between-individual estimation was obtained for state estimation by dividing the twelve activities into four high-level activity categories, suggesting that state estimation is sufficiently feasible.

Future work will focus on improving classification accuracy by exploring other classification methods such as deep learning. We also plan to investigate additional feature extraction techniques to better distinguish between overlapping activities and enhance overall system robustness. Additionally, we plan to deploy the system in real office environments to estimate user states, such as work concentration and psychological conditions.

Funding The funding was provided by Japan Society for the Promotion of Science, Grant No. JP24H00745.

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request.

References

Arnrich B, Setz C, La Marca R et al (2010) What does your chair know about your stress level? IEEE Trans Inf Technol Biomed 14(2):207–214. https://doi.org/10.1109/TITB.2009.2035498

Chang Kh, Sy Liu, Chu Hh et al (2006) The diet-aware dining table: observing dietary behaviors over a tabletop surface. Pervas Comput. https://doi.org/10.1007/11748625_23

de Macêdo TAM, Cabral ELS, Silva Castro WR et al (2020) Ergonomics and telework: a systematic review. Work 66(4):777–788. https://doi.org/10.3233/WOR-203224

Haresamudram H, Tang CI, Suh S, et al (2024) Solving the sensor-based activity recognition problem (soar): self-supervised, multi-modal recognition of activities from wearable sensors. In: Companion of the 2024 on ACM international joint conference on pervasive and ubiquitous computing, pp 1004–1007 https://doi.org/10.1145/3675094.3677562

Hernandez J, Paredes P, Roseway A, et al (2014) Under pressure: sensing stress of computer users. In: Proceedings of the SIGCHI conference on human factors in computing systems, 51–60 https://doi.org/10.1145/2556288.2557165

Ho J, Intille SS (2005) Using context-aware computing to reduce the perceived burden of interruptions from mobile devices. In: Proceedings of the SIGCHI conference on human factors in computing systems, 909–918 https://doi.org/10.1145/1054972.1055100

Inoue T, Matsusaka Y (2010) A system to recommend dishes by the real time recognition of dining activity. In: 2010 IEEE international conference on systems, man and cybernetics, 2448–2452 https://doi.org/10.1109/ICSMC.2010.5641941

Jaimes A (2006) Posture and activity silhouettes for self-reporting, interruption management, and attentive interfaces. In: Proceedings of the 11th international conference on intelligent user interfaces, 24–31 https://doi.org/10.1145/1111449.1111463

Joutou T, Yanai K (2009) A food image recognition system with multiple kernel learning. In: 2009 16th IEEE international conference on image processing, 285–288 https://doi.org/10.1109/ICIP.2009.5413400

Kern N, Schiele B, Schmidt A (2007) Recognizing context for annotating a live life recording. Pers Ubiquit Comput 11(4):251–263. https://doi.org/10.1007/s00779-006-0086-3

Kitabayashi T, Demura S, Murase H (2011) Proposal for a new body sway evaluation method. Perc Mot Skills 113(1):127–138. https://doi.org/10.2466/10.13.15.PMS.113.4.127-138

Maradana C, An M, Rasheed A, et al (2024) Human activity recognition and abnormality detection using deep learning. In: Proceedings of the 2024 6th international conference on pattern recognition and intelligent systems, pp 87–92 https://doi.org/10.1145/3689218.3689231

Matsumoto S, Wataya R, Iwai D et al (2015) Writing-state classification by ambient sedentary behavior sensing in desk work. Electr Commun Jpn 98(1):15–22. https://doi.org/10.1002/ecj.11615

Ministry of Internal Affairs and Communications, Japan (2021) Communications usage trend survey. https://www.soumu.go.jp/johotsusintokei/tsusin_riyou/data/eng_tsusin_riyou02_2021.pdf. Accessed 20 March 2024

Miyazawa K (2005) Porano no hiroba. https://www.aozora.gr.jp/cards/000081/files/1935_19925.html. Accessed 2 December 2021

Murao K, Terada T (2016) A combined-activity recognition method with accelerometers. J Inf Proc 24(3):512–521. https://doi.org/10.2197/ipsjjip.24.512

Murao K, Terada T, Yano A, et al (2012) Detecting room-to-room movement by passive infrared sensors in home environments. In: The 1st workshop on recent advances in behavior prediction and pro-active pervasive computing, 1–12

Murao K, Imai J, Terada T et al (2017) Activity recognition and user identification based on tabletop activities with load cells. J Inf Process 25:59–66. https://doi.org/10.2197/ipsjjip.25.59

Mutlu B, Krause A, Forlizzi J, et al (2007) Robust, low-cost, non-intrusive sensing and recognition of seated postures. In: Proceedings of the 20th annual ACM symposium on user interface software and technology, 149–158 https://doi.org/10.1145/1294211.1294237

OECD (2021) Teleworking in the covid-19 pandemic: trends and prospects. https://read.oecd-ilibrary.org/view/?ref=1108_1108540-p249kho0iu. Accessed 13 November 2023

Schmidt A, Strohbach M, van Laerhoven K, et al (2002) Context acquisition based on load sensing. In: UbiComp 2002: ubiquitous computing, pp 333–350 https://doi.org/10.1007/3-540-45809-3_26

Schmidt A, Strohbach M, van Laerhoven K et al (2003) Ubiquitous interaction–using surfaces in everyday environments as pointing devices. Univ Access Theor Perspect Pract Exp. https://doi.org/10.1007/3-540-36572-9_21

Tani T, Yamada S (2013) Estimating user interruptibility by measuring table-top pressure. In: CHI '13 extended abstracts on human factors in computing systems, pp 1707–1712 https://doi.org/10.1145/2468356.2468662

Wu TY, Yang XD (2022) iwood: makeable vibration sensor for interactive plywood. In: Proceedings of the 35th annual ACM symposium on user interface software and technology, pp 1–12 https://doi.org/10.1145/3526113.3545640

Yoshida T, Okazaki N, Takaki K, et al (2022) Flexel: a modular floor interface for room-scale tactile sensing. In: Proceedings of the 35th annual ACM symposium on user interface software and technology, pp 1–12 https://doi.org/10.1145/3526113.3545699

Zhang Z, Zheng H, Rempel S, et al (2020) A smart utensil for detecting food pick-up gesture and amount while eating. In: Proceedings of the 11th augmented human international conference, 1–8 https://doi.org/10.1145/3396339.3396361