Understanding Human Daily Experience Through Continuous Sensing: ETRI Lifelog Dataset 2024

Se Won Oh
Electronics and Telecommunications Research Institute
Daejeon, South Korea
sewonoh@etri.re.kr

Hyuntae Jeong
Electronics and Telecommunications Research Institute
Daejeon, South Korea
htjeong@etri.re.kr

Seungeun Chung
Electronics and Telecommunications Research Institute
Daejeon, South Korea
schung@etri.re.kr

Jeong Mook Lim
Electronics and Telecommunications Research Institute
Daejeon, South Korea
jmlim21@etri.re.kr

Kyoung Ju Noh
Electronics and Telecommunications Research Institute
Daejeon, South Korea
kjnoh@etri.re.kr

Sunkyung Lee
Electronics and Telecommunications Research Institute
Daejeon, South Korea
sklee2014@etri.re.kr

Gyuwon Jung
Electronics and Telecommunications Research Institute
Daejeon, South Korea
gwjung@etri.re.kr

***Abstract*—Improving human health and well-being requires an accurate and effective understanding of an individual's physical and mental state throughout daily life. To support this goal, we utilized smartphones, smartwatches, and sleep sensors to collect data passively and continuously for 24 hours a day, with minimal interference to participants' usual behavior, enabling us to gather quantitative data on daily behaviors and sleep activities across multiple days. Additionally, we gathered subjective self-reports of participants' fatigue, stress, and sleep quality through surveys conducted immediately before and after sleep. This comprehensive lifelog dataset is expected to provide a foundational resource for exploring meaningful insights into human daily life and lifestyle patterns, and a portion of the data has been anonymized and made publicly available for further research. In this paper, we introduce the ETRI Lifelog Dataset 2024, detailing its structure and presenting potential applications, such as using machine learning models to predict sleep quality and stress.**

***Index Terms*—Lifelog, Human Behavior, Sleep Quality, Multi-modal Data, Wearable Sensors**

I. Introduction

Human daily life consists of a complex interrelation of different activities and physiological states, spanning daytime behavior and nighttime sleep. To better understand the intricate nature of human behavior in daily life, it is important to systematically collect and analyze comprehensive lifelogs from a variety of sensor devices and user records [1], [2]. Such continuous and well-quantified data are essential for reliably assessing an individual's physical and mental states throughout the day, thereby supporting efforts to improve human health and well-being. Advances in sensor technology and data collection methods have enabled tracking of human activity and physiological signals in controlled environments or during specific tasks. In addition, mobile and wearable devices such as smartphones and smartwatches are widely used for health monitoring and behavior tracking [3]–[6].

However, despite these advances, capturing continuous and naturalistic daily experiences in real-world settings without disrupting individuals' routines remains challenging [7], [8]. Most existing studies have focused on short-term measurements or controlled laboratory environments, providing only a limited perspective on the complexity of human life. For example, although several studies have used smartphones and smartwatches to track human activity, few have comprehensively monitored a 24-hour cycle of daily life, including sleep, in natural environments over a long period of time [9]–[11].

This gap underscores the need for a dataset that seamlessly integrates continuous, passive measurements of daytime behavior and nighttime sleep patterns to reveal the underlying connections between daily activities, sleep quality, and subjective indicators of well-being such as fatigue, mood, and stress. To address this need, our study collected continuous, unobtrusive data using smartphones, smartwatches, and under-mattress sleep sensors [12]. Subjective self-reports were also taken immediately before and after sleep to capture a holistic view of daily life.

We extended previous efforts, such as the 2020 and 2023 datasets [10], [13]–[15], but further addressed some key limitations. In this new dataset, we expanded the types of sensor data collected (from 9 to 12 types) while minimizing device load and battery consumption by adjusting the data collection intervals and excluding raw acceleration data. This approach ensured a comprehensive dataset suitable for long-term observation of natural human behavior.

As with the previous ICTC, where our lifelog dataset led to diverse research outcomes [16]–[27], the ETRI Lifelog Dataset 2024 [15] introduced in this paper would provide a foundational resource for exploring meaningful insights into human routines and life patterns. For example, it can be

This work was supported by Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korean government. [25ZB1100, Core Technology Research for Self-Improving Integrated Artificial Intelligence System].

TABLE I
DEMOGRAPHIC CHARACTERISTICS OF SUBJECTS

subject_id	Gender	Age Group	Employment	BMI
id01	female	40s	Yes	23.7
id02	male	50s	Yes	22.7
id03	female	30s	No	29.1
id04	female	40s	No	19.6
id05	female	50s	Yes	26.3
id06	male	20s	Yes	27.1
id07	female	40s	Yes	26.3
id08	male	30s	Yes	29.3
id09	male	30s	Yes	25.4
id10	female	30s	Yes	26.0

used in studies to investigate how daily behaviors and sleep activities relate to subjective and physiological states. We demonstrate a real-world application of this dataset by using machine learning models to predict sleep quality, illustrating its relevance for human-centered health research.

The remainder of this paper is organized as follows: we describe the data collection methods and dataset structure, then present an application involving machine learning models for sleep quality prediction, followed by discussion and the conclusion. This study provides a robust and natural dataset that can support research on human well-being and daily life and promote the development of data-driven health solutions.

II. DATA COLLECTION AND DATASET OVERVIEW

This dataset represents a portion of the outcomes from a research project jointly conducted by Electronics and Telecommunications Research Institute (ETRI) and Chungnam National University Hospital (CNUH). It has been curated and released to promote further research and includes a total of 700 days of lifelog data collected from 10 participants who took part in experiments conducted between May and December 2024. Each subject is assigned a unique identifier (e.g., ‘subject_id’), and their demographic details are summarized in Table I. Subject ages are grouped in 10-year increments, while height and weight data are rounded for simplicity.

During daytime activities, the participants wore smartphones and smartwatches, while nighttime sleep activities were recorded using sleep sensors installed in their personal living spaces. Additionally, immediately before and after sleep, participants recorded survey inputs on fatigue, stress, and sleep quality using an ecological momentary assessment (EMA) approach.

Table II summarizes twelve data items collected using smartphones and smartwatches. The data count of each item represents the total number of data instances. Each item was collected using the participants’ Android smartphones or smartwatches, with data sampling intervals ranging from one to ten minutes. It should be noted that some data items may contain a considerable amount of missing data and noise. Since this dataset was collected in everyday life rather than

Fig. 1. Overview of the six daily metrics related to sleep health, fatigue, and stress, derived from sleep sensor data and self-reported surveys

in a controlled laboratory environment, potential noise may be present. Additionally, sensor measurements and recordings may be interrupted due to device charging or the rebooting process.

The detailed structure of individual sensor data items is summarized in Table III. Each data item is stored as an individual data file and is provided along with the subject identifier and timestamp. All timestamps are recorded based on **Korea Standard Time (KST)** and are displayed in the **YYYY-MM-DD HH:MM:SS** format (e.g., 2024-08-01 12:34:56). To protect privacy, certain data items have undergone minimal anonymization. For example, GPS latitude and longitude data are provided as relative coordinates.

In addition to the twelve data items measured from smartphones and smartwatches during daytime activities, we also derived the following six daily metrics, which are shown in Figure 1, related to sleep health, fatigue, and stress from sleep sensor data and self-reported survey records. The method used to derive these metrics is similar to that presented in our previous work [14].

- **Q1:** Overall sleep quality as perceived by a subject immediately after waking up.
- **Q2:** Physical fatigue of a subject just before sleep.
- **Q3:** Stress level of a subject just before sleep.
- **S1:** Adherence to the recommended total sleep time.
- **S2:** Adherence to the recommended sleep efficiency.
- **S3:** Adherence to the recommended sleep onset latency.

Specifically, the three survey-based metrics (i.e., Q1, Q2, and Q3) were derived from the individual’s pre- and post-sleep questionnaire responses, which were averaged across the entire experimental period. Each questionnaire was originally recorded using a 5-point Likert scale and was converted into a binary format. For instance, the first questionnaire metric (Q1) is assigned a value of **1** on days when an individual’s self-reported sleep quality exceeds their average over the experimental period, and **0** when it falls below that average. Similarly, the second and third metrics (Q2 and Q3) are assigned a value of **0** on days when the participant’s fatigue and stress levels, respectively, exceed their average, and a value of **1** when these levels are below average. In summary, the survey-based metrics are labeled as **1** when they indicate a subjectively positive outcome based on the individual participant’s self-assessment.

Meanwhile, the three sleep metrics (i.e., S1, S2, and S3), which were derived from the sleep sensor, are designed to reflect the sleep health recommendations established by the

TABLE II
SUMMARY OF DATA ITEMS COLLECTED USING SMARTPHONES AND SMARTWATCHES

Item	Data count	Sampling interval	Note
mACStatus	939,896	10-min	Indicates whether the smartphone is currently being charged
mActivity	961,062	1-min	Value calculated by the Google Activity Recognition API[a]
mAmbience	476,577	2-min	Ambient sound identification labels[b] and their respective probabilities
mBle	21,830	10-min	Bluetooth devices around individual subject
mGps	800,611	1-min	Multiple GPS coordinates measured within a single minute using the smartphone
mLight	96,258	10-min	Ambient light measured by the smartphone
mScreenStatus	939,653	1-min	Indicates whether the smartphone screen is in use
mUsageStats	45,197	10-min	Indicates which apps were used on the smartphone and for how long
mWifi	76,336	10-min	Wifi devices around individual subject
wHr	382,918	1-min	Heart rate readings recorded by the smartwatch
wLight	633,741	1-min	Ambient light measured by the smartwatch
wPedo	748,100	1-min	Step data recorded by the smartwatch

[a]https://developers.google.com/location-context/activity-recognition
[b]https://research.google.com/audioset/ontology

TABLE III
SUMMARY OF INDIVIDUAL DATA ITEM NAMES (COLUMN NAMES), DATA TYPES, AND VALUE FORMATS OR RANGES

Item	Column	Data type	Note
mACStatus	m_charging	integer	0: No, 1: Charging
mActivity	m_activity	integer	0: in_vehicle, 1: on_bicycle, 2: on_foot, 3: still, 4: unknown, 5: tilting, 7: walking, 8: running
mAmbience	m_ambience	object	List of ambient sound labels and their respective probabilities
mBle	m_ble	object	List of bluetooth device address, device_class, and rssi
mGps	m_gps	object	List of (altitude, latitude, longitude, speed)
mLight	m_light	float	Ambient light in lx unit
mScreenStatus	m_screen_use	integer	0: No, 1: Using screen
mUsageStats	m_usage_stats	object	List of app names and their respective usage times (in milliseconds unit)
mWifi	m_wifi	object	List of base station ID(bssid) and rssi
wHr	heart_rate	object	List of heart rate recordings
wLight	w_light	float	Ambient light in lx unit
wPedo	burned_calories	float	Number of calories
	distance	float	Distance in meters
	speed	float	Speed in km/h unit
	step	integer	Number of steps
	step_frequency	float	Step frequency in a minute

National Sleep Foundation [28]. These metrics assign a value of **0** for sleep records that do not meet the recommended guidelines. In contrast, sleep records that satisfy the recommendations are assigned a value of **1** for S2 and S3, which are binary metrics. Notably, unlike the other metrics, the sleep metric for total sleep time (S1) can take on three distinct values(e.g., **0**: Not recommended; **1**: May be appropriate; **2**: Recommended).

III. BASELINE MACHINE LEARNING MODEL

In this section, we present an example of a baseline machine learning model to demonstrate the types of features that can be extracted and the classification performance achievable when these features are input into a model to classify four target metrics: Q1, Q2, Q3, and S3. Figure 2 illustrates the number of instances labeled as **0** and **1** for these four metrics, highlighting potential class imbalance.

Out of the total 700 days of Lifelog data, 450 days were allocated for training a learning model, while the remaining 250 days were used for evaluation. Since the data were collected from human participants and behavioral changes may occur over extended experimental periods, the test data were not limited to the final phase of participation. Instead, the test data were configured to include both the middle and later stages of each subject's participation period, thereby better reflecting temporal variations in behavior. It is worth noting that a chi-squared test revealed no significant differences in the distribution of each target metric between the training and test data splits ($p > 0.2$). Figure 3 shows the temporal

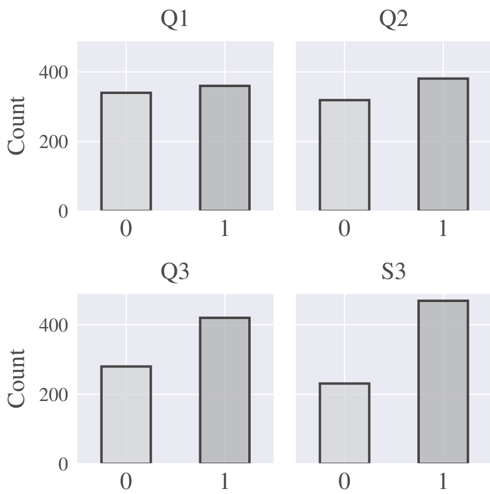

Fig. 2. Label distribution (0 and 1) for the four target metrics

distribution of training and test data points for each subject, highlighting the non-contiguous and personalized split across different dates. The discontinuity in daily records is due to the prior exclusion of days that either failed to meet the minimum metric generation criteria (e.g., less than three hours spent in bed) or exhibited excessive missing data.

We derived features from five data sources: wPedo, wHr, wLight, mLight, and mUsage. Given that some data items may be missing in the original dataset, we decided to use descriptive features at fixed time intervals to cope with information loss and reduce the computational burden on the learning model. For wPedo, wHr, wLight, and mLight, we segmented each day into seven time zones (i.e., 00h-06h, 06h-09h, 09h-12h, 12h-15h, 15h-18h, 18h-21h, and 21h-24h) and computed representative statistical values such as the mean, sum, and standard deviation.

- wPedo: Mean and total step count per time zone
- wHr: Mean heart rate per time zone and proportion of high heart rate values
- wLight, mLight: Mean and standard deviation of light intensity per time zone after applying log10 transformation

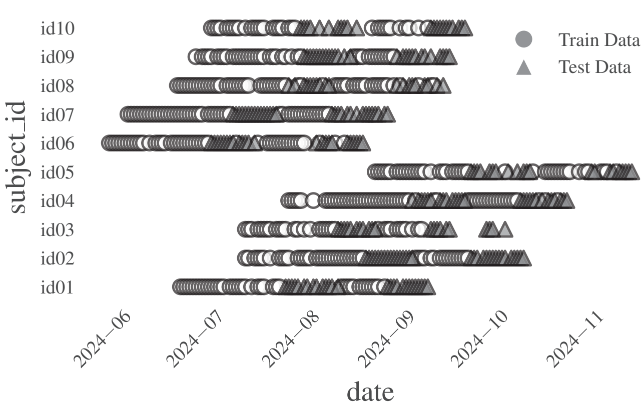

Fig. 3. Temporal distribution of training (triangle) and test (circle) data across individual subjects

For mUsage, we categorized the most frequently used mobile apps into three broad categories (i.e., System, Social, and Hobby) and computed the total daily usage time in minutes for each category. In addition, four demographic attributes for each subject (i.e., gender, age group [whether aged 40 or older], employment status, and BMI), along with day-of-week information, were included as input features (marked as 'DW') for model training.

To assess the baseline model's performance with the extracted features, we trained classification models using Python 3.12, LightGBM (version 4.6), and scikit-learn (version 1.6). All models were trained using default parameter settings provided by the LightGBM library. Table IV presents the classification performance of the baseline model for each target metric, measured by macro F1 score. Bolded values indicate the feature combinations that achieved the highest performance in our experiments.

IV. Discussion

The ETRI Lifelog dataset 2024 included more participants (increased from 4 to 10) and a larger number of daily instances (from 220 to 700 days) compared to the 2023 dataset. However, among the seven target metrics defined in 2023 [14], some were revised in 2024, resulting in a total of six daily metrics. For instance, the Q2 metric, which represented the emotional state just before sleep in 2023, was replaced in 2024 to indicate the fatigue level before sleep. Similarly, the S1 metric for total sleep time, previously a binary classification, was modified into a three-class classification in 2024.

Figure 4 presents the Spearman correlations among six metrics for the ETRI Lifelog Dataset 2024. Only statistically

TABLE IV
Performance comparison across different feature combinations

Features (n)	**Q1**	**Q2**	**Q3**	**S3**
wPedo (14)	0.549	0.524	0.563	0.537
wPedo+DW (19)	**0.619**	0.537	0.647	**0.651**
wHr (14)	0.548	0.492	0.587	0.526
wHr+DW (19)	0.527	0.568	0.577	0.539
wLight (14)	0.513	0.563	0.540	0.524
wLight+DW (19)	0.557	0.565	0.604	0.594
mLight (14)	0.477	0.557	0.562	0.507
mLight+DW (19)	0.579	0.574	0.592	0.641
mUsageStats (3)	0.543	0.568	0.497	0.554
mUsageStats+DW (8)	0.562	**0.634**	0.623	0.610
wPedo+mUsageStats (17)	0.563	0.532	0.643	0.544
wPedo+mUsageStats+DW (22)	0.563	0.596	**0.674**	0.575
All (59)	0.531	0.552	0.635	0.603
All+DW (64)	0.570	0.569	0.597	0.594

Note. DW means five features about demographics and weekday information.

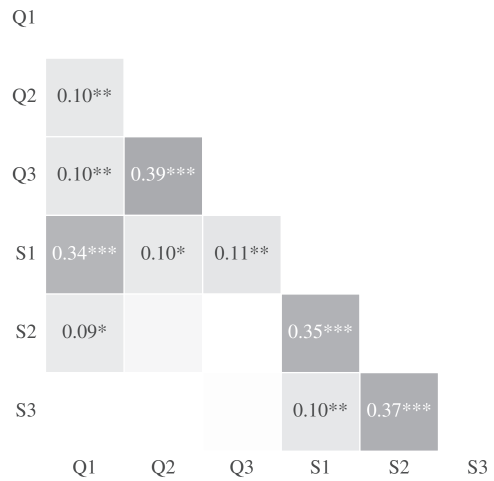

Fig. 4. Spearman correlation matrix among six daily metrics. Asterisks indicate significance levels: p<0.05 (*), p<0.01 (**), and p<0.001 (***).

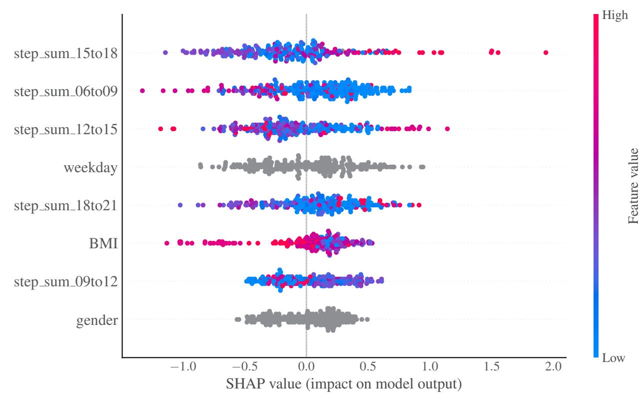

Fig. 5. SHAP values for the baseline model predicting Q1 using wPedo+DW features

significant coefficients ($p<0.05$) are shown in the figure. The results indicate that the metrics exhibit weak positive correlations, either directly or indirectly. For example, Q1 is not directly correlated with S3, but it shows a significant correlation with S1, which in turn correlates with S2, and S2 with S3, which implies an indirect association between Q1 and S3. This suggests that the perceived quality of sleep is associated with various behavioral outcomes during sleep. A positive association is also observed between Q2 and Q3. This observation indicates that physical fatigue and stress levels before sleep are positively related.

An analysis of the modeling results presented in Table IV reveals that the baseline models generally performed better when input features were selectively chosen for each metric, rather than when all available features were used. This finding suggests that machine learning models benefit from a more refined and task-specific feature selection process. These results are consistent with those reported in a previous study conducted last year. For instance, in the cases of Q1 and S3, the highest performance was achieved using features related to step counts in 3-hour time units, while for Q2 and Q3, features based on daily mobile app usage yielded superior results. Additionally, across all baseline models, the inclusion of demographic and weekday features consistently enhanced classification performance.

Figure 5 presents SHAP value analysis [29] for the classification model using 'wPedo+DW' features, which showed the best performance in predicting the Q1 metric. The plot displays the top 8 features ranked by their impact on the model's output. The most influential features, in order, are step count from 15:00 to 18:00, step count from 06:00 to 09:00, step count from 12:00 to 15:00, and whether the day is a weekday. Notably, higher step counts between 15:00 and 18:00 contribute positively to predicting Q1 = **1**, while higher step counts between 06:00 and 09:00 contribute negatively.

Figure 6 illustrates the SHAP values for the weekday feature across different days of the week. The SHAP values tend to be higher on Fridays and lower on Sundays and Tuesdays, which aligns with general intuition. This suggests that, in addition to sensor-derived features, contextual factors related to participants' daily environments play a role in model predictions. It also helps explain the performance improvement observed in all models when 'DW' features (i.e., demographics and weekday information) were included as input, as previously shown in Table IV.

V. Conclusion

In this paper, we detailed the main components and characteristics of the ETRI Lifelog Dataset 2024. We also presented a baseline machine learning model for sleep quality prediction along with an example analysis of feature importance using SHAP values. This dataset has been made publicly available for research and has served as a key resource for the fourth Human Understanding AI Paper Challenge, which was held in conjunction with ICTC 2025. We anticipate that this new dataset will likewase provide a valuable foundation for a wide range of studies, including those on physical activity, sleep quality, stress patterns, and broader aspects of daily human life.

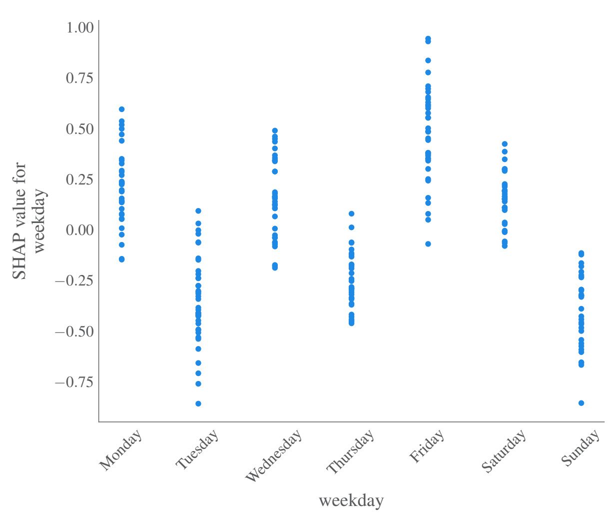

Fig. 6. SHAP values by day of the week (weekday feature)

ACKNOWLEDGMENT

This work was supported by Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korean government. [25ZB1100, Core Technology Research for Self-Improving Integrated Artificial Intelligence System].

REFERENCES

[1] Chan, Shing, et al., "CAPTURE-24: A large dataset of wrist-worn activity tracker data collected in the wild for human activity recognition," Scientific Data, vol. 11(1), 2024.

[2] Nepal, Subigya, et al., "Capturing the college experience: a four-year mobile sensing study of mental health, resilience and behavior of college students during the pandemic," Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies, vol. 8(1), pp. 1–37, 2024.

[3] Huhn, Sophie, et al., "The impact of wearable technologies in health research: scoping review," JMIR mHealth and uHealth, vol. 10(1), 2022.

[4] Niknejad, N., Ismail, W. B., Mardani, A., Liao, H., and Ghani, I., "A comprehensive overview of smart wearables: The state of the art literature, recent advances, and future challenges," Engineering Applications of Artificial Intelligence, vol. 90, 2020.

[5] Peake, J. M., Kerr, G., and Sullivan, J. P., "A critical review of consumer wearables, mobile applications, and equipment for providing biofeedback, monitoring stress, and sleep in physically active populations," Frontiers in physiology, vol. 9, 2018.

[6] Hardjianto, Mardi, et al., "A graph neural network model application in point cloud structure for prolonged sitting detection system based on smartphone sensor data," ETRI Journal, vol. 47(2), pp. 290–302, 2025.

[7] Hicks, Jennifer L., et al., "Best practices for analyzing large-scale health data from wearables and smartphone apps," NPJ digital medicine vol. 2(1), 2019.

[8] Vaizman, Y., Weibel, N., and Lanckriet, G., "Context recognition in-the-wild: Unified model for multi-modal sensors and multi-label classification," Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 1(4), pp. 1–22, 2018.

[9] Yfantidou, Sofia, et al., "LifeSnaps, a 4-month multi-modal dataset capturing unobtrusive snapshots of our lives in the wild," Scientific Data, vol. 9(1), 2022.

[10] Chung, S., et al., "Real-world multimodal lifelog dataset for human behavior study," ETRI Journal, vol. 44(3), pp. 426–437, 2022.

[11] Wang, Rui, et al., "Tracking depression dynamics in college students using mobile phone and wearable sensing," Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 2(1), pp. 1–26, 2018.

[12] Lim, J., Chung, S., Oh, S., Noh, K. and Jeong, H., "Digital Phenotype Collection System Utilizing Smart Devices," ICTC 2023, pp. 1709–1711, Oct. 2023.

[13] ETRI, "ETRI Lifelog Dataset 2020," 2021. [Online]. Available: https://nanum.etri.re.kr/share/schung/ETRILifelogDataset2020

[14] Oh, S., Jeong, H., Chung, S., Lim, J., and Noh, K., "Sensor-based multi-label dataset analysis challenge: predicting sleep quality and emotional states in daily life," ICTC 2024, pp. 771–775, Oct. 2024.

[15] ETRI, "ETRI Lifelog Dataset 2024," 2025. [Online]. Available: https://nanum.etri.re.kr/share/human/ETRILifelogDataset2024

[16] Na, Y., Oh, S., Ko, S., and Lee, H., "Pixleepflow: A pixel-based lifelog framework for predicting sleep quality and stress level," ICTC2024, pp. 810–815, Oct. 2024.

[17] Kim, J., Ma, M., Choi, E., Cho, K., and Lee, C., "Tram: Enhancing user sleep prediction with transformer-based multivariate time series modeling and machine learning ensembles," ICTC2024, pp. 822–827, Oct. 2024.

[18] Kim, S., Kim, J., Shin, H., and Kim, S., "Predicting sleep quality using lifelog data with deep learning techniques," ICTC2024, pp. 776–780, Oct. 2024.

[19] Lee, T., Ha, S., and Lee, H., "Predicting mental health using lifelog data: Application of median resampling and data augmentation techniques," ICTC2024, pp. 793–797, Oct. 2024.

[20] Ham, J., Ha, Y., Yoo, K., and Baek, J., "Human understanding from lifelog data via mobile sensors using a gru network with attention mechanism," ICTC2024, pp. 834–839, Oct. 2024.

[21] Lee, J., Yu, S., Kim, D., and Choi, K., "Multi-patching: Life-log classification with the reconstructed representation of multivariate time series," ICTC2024, pp. pages 798–803, Oct. 2024.

[22] Cho, Y., Kwon, N., Yoon, B., and Kim, J., "Multi-modal sensor fusion for predicting sleep and stress patterns," ICTC2024, pp. 804–809, Oct. 2024.

[23] Kang, S., Kim, J., and Kim, N., "Research on memory-efficient approach to sleep quality estimation for edge devices," ICTC2024, pp. 781–786, Oct. 2024.

[24] Kim, S., Ahn, S., Kim, M., Lee, M., and Jeong, D., "Human everyday experience metric recognition based on classifier chains using heart rate," ICTC2024, pp. 828–833. IEEE,2024.

[25] Park, J., Kim, Y., Song, J., and Lee, H., "Contextual health state inference from lifelog data using llm," ICTC2024, pp. 816–821, Oct. 2024.

[26] Ro, J. and Yoon, Y., "Fill in the gap: Two-stage prediction modeling framework for understanding human biosensor data," ICTC2024, pp. 787–792, Oct. 2024.

[27] Seong, H., Ahn, S., Lee, H., and Lee, K., "Efficient feature engineering and lightweight ai models for predicting lifestyle indicators using lifelog data," ICTC2024, pp. 1003–1008, Oct. 2024.

[28] National Sleep Foundation, "NSF's Guidelines," 2024. [Online]. Available: https://www.thensf.org/guidelines

[29] Lundberg, S.M., and Lee, S.-I., "A unified approach to interpreting model predictions," Advances in neural information processing systems 30, 2017.