CallSense: Sensing Call Engagement in Privacy-Constrained Call Center Environment with Commodity Smartwatches

Bonwoo Koo
Graduate School of Data Science
Korea Advanced Institute of Science and Technology
Daejeon, Republic of Korea
bonwoo.koo@kaist.ac.kr

Junmo Lee
School of Computing
Korea Advanced Institute of Science and Technology
Daejeon, Republic of Korea
junmo.lee@kaist.ac.kr

Uichin Lee
AI Computing Department
Korea Advanced Institute of Science and Technology
Daejeon, Republic of Korea
uclee@kaist.ac.kr

Abstract

Call center agents face chronic emotional labor, and just-in-time well-being interventions should avoid interrupting active calls. We present *CallSense*, a smartwatch-based approach for sensing call engagement using accelerometer and photoplethysmography signals without audio or call-log access. We collected four weeks of in-the-wild data from 19 call center agents and evaluated user-independent and user-dependent classifiers for distinguishing call engagement and between-call periods. The best AUROC was 0.764 for unseen users and 0.854 for seen users, with accelerometer signals accounting for most of the discriminative performance. These results support the feasibility of call engagement sensing from commodity wearables as a privacy-aware basis for well-being support.

CCS Concepts

• **Human-centered computing → Ubiquitous and mobile computing systems and tools**.

Keywords

wearable sensing; smartwatch; emotional labor; well-being support

ACM Reference Format:
Bonwoo Koo, Junmo Lee, and Uichin Lee. 2026. CallSense: Sensing Call Engagement in Privacy-Constrained Call Center Environment with Commodity Smartwatches. In *Companion of the 2026 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp Companion '26), October 11–15, 2026, Shanghai, China.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3798063.3841762

1 Introduction

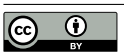
Emotion workers such as call center agents regulate their emotional expressions according to organizational display rules [15], and the chronic emotional labor inherent in this role places them at heightened risk of burnout, anxiety, and depression [21]. To support their well-being, prior work has explored passive sensing to detect stress in this population [17, 21]. Stress detection, however, is only useful when it can be translated into just-in-time (JIT) interventions that help workers recover or self-regulate [19]. A critical caveat is that intervention timing itself matters: delivering a prompt at an inopportune moment can disrupt task execution and impose additional cognitive cost rather than relieve it [1]. For call center agents, this implies that an intervention triggered during a call engagement (CE) could exacerbate stress, and that only the between-call (BC) periods offer safe windows for support. Reliable detection of the CE/BC cycle is therefore a prerequisite for any deployable JIT intervention in this population.

The most reliable source of CE/BC state is the audio data and internal call-log systems. Recent work has demonstrated that task-log features dominate after-call stress detection in call centers [21], underscoring the operational signal embedded in call records. In practice, however, call centers operate under strict security policies that restrict both the repurposing of internal call logs and recorded audio, even when processed locally on a worker's device. This calls for a worker-owned CE sensing system that does not rely on data sources controlled by an organization.

Prior works have shown that passive physiological signals can distinguish work and break states of knowledge workers [7, 11, 12, 26]. Schaule et al. [26] used commodity wearables to detect cognitive load and identify opportune interruption moments in a controlled office setting, and Di Lascio et al. [12] demonstrated in-the-wild recognition of breaks and work activities from wrist-worn physiological signals (EDA, BVP, ACC) along with phone and laptop usage. Among these, the accelerometer and PPG channels available on commodity smartwatches make them plausible candidates for distinguishing CE and BC without external infrastructure.

However, the call center environment differs from the knowledge worker contexts in which these signals have been validated in two important ways. First, the temporal granularity of state transitions is far finer: while knowledge workers engage in continuous work over hourly timescales, call agents alternate between CE and BC on a per-minute scale to handle a high rate of incoming calls [16]. The five-minute feature windows adopted in prior physiological work-break classification studies [11, 12, 26] are poorly matched to this rapid alternation. Second, the behavioral granularity between states is far finer: while knowledge workers often leave the workstation during breaks, call agents typically remain at the same desk during BC to stay available for the next call. Call center agents thus represent a distinct case of interaction-bound service work that remains comparatively underexplored in wearable sensing research.

To address this gap, we present *CallSense*, a smartwatch-based system that senses call engagement to identify safe windows for well-being support, using only the accelerometer and PPG signals, without audio or call-log access (Figure 1). We collected four weeks of in-the-wild data from 19 call center agents at a metropolitan call



This work is licensed under a Creative Commons Attribution 4.0 International License.
UbiComp Companion '26, Shanghai, China.
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2533-3/2026/10
https://doi.org/10.1145/3798063.3841762

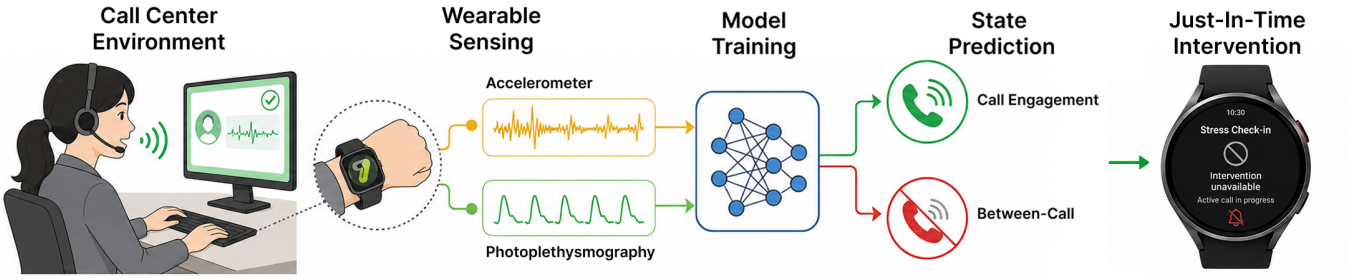


Figure 1: Overview of *CallSense*. Wearable accelerometer and photoplethysmography signals are used to classify call engagement (CE) and between-call (BC) states, identifying safe windows for just-in-time intervention in privacy-constrained call center environments.

center, and analyzed which features are most strongly associated with the CE/BC distinction.

2 Methodology

2.1 Data

We collected data over four weeks from 19 call center agents (2 male, 17 female; age $M = 38.89$, $SD = 5.81$) employed at a metropolitan call center in South Korea. Although our sample is female-skewed, this distribution is consistent with the South Korean call center workforce, where prior studies report that women make up over 90% of agents [10]. Each participant wore a Samsung Galaxy Watch 7 during their shift, which recorded a triaxial accelerometer signal at 25 Hz, a green PPG signal at 25 Hz, and SDK-derived heart rate (HR)[1] at 1 Hz. In parallel, call start and end timestamps were extracted from the call center's internal call logs to label the ground-truth data. All participants provided informed consent prior to participation, and this study was approved by the Institutional Review Board.

2.2 Call Engagement Labeling

Using each participant's call logs, we labeled every second of the wearable data stream as either a CE or BC segment. A CE segment was defined by the logged call start and end timestamps, and the immediately following idle interval was treated as the corresponding BC segment. To remove unreliable segments, we discarded any segment shorter than 60 s and further removed segments whose duration exceeded the 95th percentile of that participant's segment-duration distribution. After filtering, the final dataset comprised 4,305 CE segments ($M = 143.42$ s, $Mdn = 128$ s, $SD = 64.59$ s) and 5,087 BC segments ($M = 349.77$ s, $Mdn = 316$ s, $SD = 185.16$ s).

3 Experiment

3.1 Data Preprocessing

Within each CE or BC segment, raw sensor data were segmented into 60 s windows with 50% overlap, discarding trailing windows under 60 s. We then applied gating based on SDK-derived heart rate, retaining windows with at least 90% valid samples (hrStatus = 1, 40–200 BPM). Missing HR samples were linearly interpolated. Since CE segments last ~2 minutes on average while BC segments last ~6 minutes, partitioning them into 60 s windows yields 13,488 CE and 52,386 BC windows, an ~1:4 class ratio.

For the ACC signal, we applied a Hampel filter to suppress impulse outliers [22] and a 4th-order Butterworth low-pass filter (cut-off 3 Hz) [2]. For the PPG signal, the irregularly sampled data were first interpolated using PCHIP [5], after which a 4th-order Butterworth bandpass filter (0.5–4 Hz) [13] was applied. The signal was then linearly detrended and z-score normalized.

Table 1: Feature set used by the ML classifiers (54 features).

ACC features {x, y, z, magnitude} (28)	
Time-domain (7)	mean, std, range, slope, mean absolute slope, first-difference mean, first-difference std
PPG features (26)	
Time-domain (2)	mean, std
Heart Rate (4)	median, min, max, std of HR
Heart Rate Variability (10)	NN mean, SDNN, SDSD, RMSSD, NN count, NN median, NN min, NN max, NN std, NN coefficient of variation
Pulse Morphology (10)	pulse amplitude mean / std, inter-peak interval mean / std, adjacent-peak amplitude-difference mean / std, adjacent-trough amplitude-difference mean / std, pulse area mean / std

3.2 Feature Extraction

For the machine learning (ML) classifiers with feature-level input, we extracted 54 handcrafted features per 60 s window across the two modalities (Table 1). The time-domain and heart rate-related features were adopted from prior work on physiological-signal-based work–break detection, stress detection, and talking detection [3, 12, 21]. In contrast, the DL model directly consumes the preprocessed ACC (x, y, z, magnitude) and PPG signals as input, without explicit feature extraction.

[1] HR was computed on-device via PPG signals and the Samsung Health Sensor SDK.

Table 2: Call Engagement Detection. Best AUROC per modality in bold. Threshold tuned on validation.

	ACC				PPG				ACC + PPG			
Model	Acc%	Sens%	Spec%	AUROC	Acc%	Sens%	Spec%	AUROC	Acc%	Sens%	Spec%	AUROC
User-independent validation												
Random Predictor	67.6	20.8	79.7	0.502	67.6	20.8	79.7	0.502	67.6	20.8	79.7	0.502
XGB	65.6	75.7	62.8	0.752	42.4	81.4	32.4	0.598	65.6	72.5	63.6	0.738
CNN-BiLSTM	57.7	79.8	51.9	0.697	38.6	82.2	27.5	0.566	62.6	72.0	60.2	0.712
TabPFN	66.9	73.0	65.1	**0.754**	43.8	81.5	34.1	**0.608**	68.7	70.7	68.0	**0.764**
User-dependent validation												
Random Predictor	67.8	19.9	78.3	0.491	67.8	19.9	78.3	0.491	67.8	19.9	78.3	0.491
XGB	77.1	67.7	79.0	0.828	47.8	78.5	40.4	0.652	77.1	66.8	79.1	0.812
CNN-BiLSTM	68.5	72.7	67.4	0.773	36.1	85.2	25.2	0.610	70.2	69.0	70.0	0.785
TabPFN	80.0	72.9	81.5	**0.854**	49.9	79.5	42.7	**0.673**	78.6	72.9	79.9	**0.848**

3.3 Models

To establish a chance-level reference, we include a random baseline, a stratified random predictor that samples labels in proportion to the training-set class frequencies while ignoring the input. For the ML model, we use XGBoost (XGB), a gradient-boosted decision-tree classifier widely applied to tabular physiological features [12]. For the DL model, we use CNN-BiLSTM, a hybrid convolutional-recurrent backbone commonly used to encode physiological signals in human-activity and emotion recognition [24]. Each window passes through a 1D convolutional front-end, a bidirectional LSTM whose hidden states are mean-pooled into a fixed-length embedding, and a fully connected layer that maps this embedding to the class logit. To better account for our small and class-imbalanced dataset, we further adopt TabPFN [18]. TabPFN is a tabular foundation model that performs in-context learning via a transformer pre-trained on large-scale synthetic tabular tasks, producing well-calibrated predictions on small datasets without per-task hyperparameter tuning. For multimodal fusion, we use a classifier ensemble for XGB, sensor-based late fusion for CNN-BiLSTM, and feature-level fusion for TabPFN [11].

3.4 Evaluation

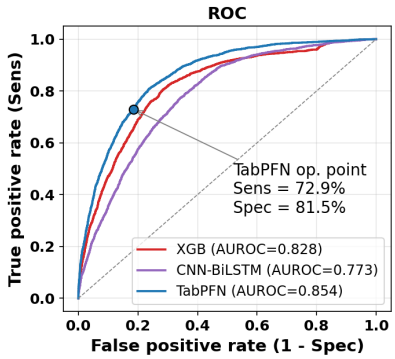
Models were evaluated under 5-fold nested cross-validation for user-independent evaluation [8], and under a chronological 80/20 within-participant split for user-dependent evaluation. Metrics were computed per fold or per participant and macro-averaged. XGB and CNN-BiLSTM were tuned with Optuna over an inner GroupKFold cross-validation. We report accuracy and AUROC as overall performance measures, with AUROC as the primary metric given the class imbalance. Since interrupting an active call and missing a between-call window carry asymmetric costs, we additionally report sensitivity and specificity, which expose the two error directions separately. Sensitivity reflects how often active calls are correctly left uninterrupted, and specificity reflects how often between-call windows available for support are recovered. We selected the threshold on out-of-fold validation predictions and applied it to the held-out test set, using a single pooled threshold for user-independent validation and a per-participant threshold for user-dependent validation.

4 Results

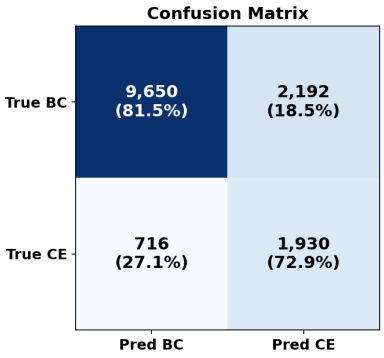
We evaluated *CallSense* under both user-independent and user-dependent validation to assess feasibility for unseen and seen users. All learned models exceeded the random baseline, which stayed at chance level with an AUROC around 0.50, confirming that *CallSense* captures a real physiological signal rather than exploiting the class prior (Table 2). ACC-based models performed best across both schemes. In user-independent validation, TabPFN on ACC+PPG reached the highest AUROC of 0.764, with TabPFN on ACC alone close behind at 0.754. In user-dependent validation, TabPFN on ACC reached the highest AUROC of 0.854. While PPG-only models still separated the two states above chance, multimodal fusion remained on par with ACC alone. For TabPFN on ACC, sensitivity and specificity reached 72.9% and 81.5%, indicating that most active calls were correctly left uninterrupted and most between-call windows available for support were recovered.

5 Discussion

The learned models improved sensitivity and AUROC over the random baseline across both validation schemes, indicating that they captured physiological differences between call engagement and between-call windows. Their best performance came from the accelerometer modality, identifying accelerometer signals as the most informative for this distinction. Among the learned models, XGB and TabPFN performed comparably, whereas CNN-BiLSTM was less competitive. This discrepancy is likely attributable to dataset scale, as training a deep model from scratch is challenging with limited in-the-wild data. TabPFN instead relies on a prior learned from synthetic data, which is better suited to this regime and yielded the highest performance. The moderate results from conventional models, further improved by TabPFN's prior, suggest that call engagement sensing from worker-owned wearables is feasible and could strengthen with larger data.



(a) ROC curves for XGB, CNN-BiLSTM, and TabPFN. TabPFN achieves the highest AUROC of 0.854. The marked operating point corresponds to a sensitivity of 72.9% and a specificity of 81.5%.



(b) Confusion matrix for TabPFN at the optimal operating point.

Figure 2: Best model performance under user-dependent validation on the ACC modality.

To identify features most relevant to CE sensing, we analyzed feature importance of the interpretable tree-based XGB model using the mean decrease in impurity method [6]. Two accelerometer features dominated: accelerometer z-axis mean and accelerometer x-axis absolute slope mean. The higher acc-z mean during CE is consistent with a palm-down wrist posture associated with keyboard typing [27], whereas the greater acc-x absolute slope during BC reflects more dynamic forearm movements without this postural constraint [14]. These patterns match prior in-the-wild observations of greater movement variability during breaks than work [12]. In contrast, the most important PPG feature, PPG NN mean, an indirect measure of heart rate, differed only marginally between CE and BC. This accounts for the weak PPG-only models and the limited gain from fusion, and aligns with prior wearable studies where movement- and task-related features dominate over cardiac signals in real-world conditions [20, 21].

For TabPFN (Figure 2a), a sensitivity of 72.9% and specificity of 81.5% translate directly into intervention costs: about 27% of active call windows would be interrupted, while about 19% of between-call windows available for support would be missed (Figure 2b). Since interrupting a call is the more costly error, a deployed system should prioritize sensitivity and adopt an operating point that lowers the mid-call intervention rate, even at the expense of some missed opportunities. The user-dependent setting requires per-user CE and BC labels. A brief enrollment phase could supply these from call logs or ecological momentary assessment (EMA), after which the system operates label-free. The high AUROC of 0.854 leaves room to move toward such a sensitivity-oriented operating point while preserving the ability to separate CE from BC. Tuning this operating point to the asymmetric cost of mid-call interruption is thus a concrete step toward reliably identifying safe windows for intervention in call center settings.

6 Limitations and Future Work

The clearest limitation of our system is the marginal contribution of PPG. Unlike in stress detection or talking detection, where cardiac features can be discriminative on their own, CE sensing is a broader problem. Agents are not necessarily stressed during every call, nor relaxed between calls, so cardiac signals alone cannot mark the CE/BC boundary. Prior mobile sensing work has used respiration to detect conversation [4]. Respiration can also be derived from PPG [9], which could capture information the accelerometer does not. Future work can explore respiration-related PPG features to better detect CE periods.

Another limitation is dataset scale and scope. Our sample is small and drawn from a single call center in one country, which may limit both the robustness of the learned models and their generalization. Larger, multi-site data would allow a firmer assessment, and wearable foundation models pretrained on physiological signals [23, 25, 28] can offer a complementary path. Such models could inject modality-specific prior knowledge that a small in-the-wild dataset cannot provide on its own.

7 Conclusion

We presented *CallSense*, a smartwatch-based call engagement sensing system that uses only accelerometer and PPG signals, without audio or call-log access. Using four weeks of in-the-wild data from 19 call center agents, TabPFN reached an AUROC of 0.764 for unseen users and 0.854 for seen users. Postural accelerometer signals emerged as the dominant indicators of CE, whereas PPG-derived cardiac signals differed only marginally. These results demonstrate the feasibility of call engagement sensing from worker-owned wearables. By identifying safe windows for support without accessing audio or call logs, *CallSense* offers a privacy-aware, sensing-based approach to provide the prerequisites for well-being interventions aimed at emotion workers.

8 Acknowledgments

This research was supported by the Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) (RS-2022-II220064) and by the National Research Foundation of Korea (NRF) funded by the Korean government (MSIT) (RS-2022-NR068758).

References

[1] Piotr D. Adamczyk and Brian P. Bailey. 2004. If not now, when? The effects of interruption at different moments within task execution. In *Proceedings of*

the SIGCHI Conference on Human Factors in Computing Systems (CHI '04). ACM, 271–278. doi:10.1145/985692.985727

[2] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, Jorge Luis Reyes-Ortiz, et al. 2013. A public domain dataset for human activity recognition using smartphones.. In *Esann*, Vol. 3. 3–4.

[3] Erick Javier Argüello-Prada, María Alejandra Dávalos Cantín, and Juan Camilo Victoria. 2023. A photoplethysmography-based system for talking detection in bedridden patients. *Biomedical Signal Processing and Control* 81 (2023), 104477.

[4] Rummana Bari, Roy J Adams, Md Mahbubur Rahman, Megan Battles Parsons, Eugene H Buder, and Santosh Kumar. 2018. rconverse: Moment by moment conversation detection using a mobile respiration sensor. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 2, 1 (2018), 1–27.

[5] Mouna Benchekroun, Baptiste Chevallier, Vincent Zalc, Dan Istrate, Dominique Lenne, and Nicolas Vera. 2023. The impact of missing data on heart rate variability features: A comparative study of interpolation methods for ambulatory health monitoring. *IRBM* 44, 4 (2023), 100776.

[6] Leo Breiman. 2001. Random forests. *Machine learning* 45, 1 (2001), 5–32. doi:10.1023/A:1010933404324

[7] Scott Allen Cambo, Daniel Avrahami, and Matthew L. Lee. 2017. BreakSense: Combining Physiological and Location Sensing to Promote Mobility during Work-Breaks. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, 3595–3607. doi:10.1145/3025453.3026021

[8] Gavin C Cawley and Nicola LC Talbot. 2010. On over-fitting in model selection and subsequent selection bias in performance evaluation. *The Journal of Machine Learning Research* 11 (2010), 2079–2107.

[9] Peter H. Charlton, Drew A. Birrenkott, Timothy Bonnici, Marco A. F. Pimentel, Alistair E. W. Johnson, Jordi Alastruey, Lionel Tarassenko, Peter J. Watkinson, Richard Beale, and David A. Clifton. 2018. Breathing Rate Estimation From the Electrocardiogram and Photoplethysmogram: A Review. *IEEE Reviews in Biomedical Engineering* 11 (2018), 2–20. doi:10.1109/RBME.2017.2763681

[10] Seong-Sik Cho, Hyunjoo Kim, JinWoo Lee, Sinye Lim, and Woo Chul Jeong. 2019. Combined exposure of emotional labor and job insecurity on depressive symptoms among female call-center workers: A cross-sectional study. *Medicine* 98, 12 (2019), e14894.

[11] Elena Di Lascio, Shkurta Gashi, Maike E Debus, and Silvia Santini. 2021. Automatic recognition of flow during work activities using context and physiological signals. In *2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 1–8.

[12] Elena Di Lascio, Shkurta Gashi, Juan Sebastian Hidalgo, Beatrice Nale, Maike E Debus, and Silvia Santini. 2020. A multi-sensor approach to automatically recognize breaks and work activities of knowledge workers in academia. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 3 (2020), 1–20.

[13] Mohamed Elgendi. 2012. On the analysis of fingertip photoplethysmogram signals. *Current cardiology reviews* 8, 1 (2012), 14–25.

[14] Martin Gjoreski, Hristijan Gjoreski, Mitja Luštrek, and Matjaž Gams. 2016. How accurately can your wrist device recognize daily activities and detect falls? *Sensors* 16, 6 (2016), 800.

[15] Alicia A. Grandey. 2000. Emotional regulation in the workplace: A new way to conceptualize emotional labor. *Journal of Occupational Health Psychology* 5, 1 (2000), 95–110. doi:10.1037/1076-8998.5.1.95

[16] Simone Grebner, Norbert Semmer, Luca Lo Faso, Stephan Gut, Wolfgang Kälin, and Achim Elfering. 2003. Working conditions, well-being, and job-related attitudes among call centre agents. *European Journal of Work and Organizational Psychology* 12, 4 (2003), 341–365.

[17] Javier Hernandez, Robert R. Morris, and Rosalind W. Picard. 2011. Call Center Stress Recognition with Person-Specific Models. In *Affective Computing and Intelligent Interaction (ACII 2011) (Lecture Notes in Computer Science, Vol. 6974)*. Springer, 125–134. doi:10.1007/978-3-642-24600-5_16

[18] Noah Hollmann, Samuel Müller, Lennart Purucker, Arjun Krishnakumar, Max Körfer, Shi Bin Hoo, Robin Tibor Schirrmeister, and Frank Hutter. 2025. Accurate predictions on small data with a tabular foundation model. *Nature* 637, 8045 (2025), 319–326. doi:10.1038/s41586-024-08328-6

[19] Esther Howe, Jina Suh, Mehrab Bin Morshed, Daniel McDuff, Kael Rowan, Javier Hernandez, Marah Ihab Abdin, Gonzalo Ramos, Tracy Tran, and Mary P. Czerwinski. 2022. Design of Digital Workplace Stress-Reduction Intervention Systems: Effects of Intervention Type and Timing. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. ACM, 1–16. doi:10.1145/3491102.3502027

[20] Murat Kucukosmanoglu, Sarah Conklin, Kanika Bansal, Sena Kaya, Yumna Anwar, Quang Dang, Golshan Kargosha, Justin Brooks, Cody Feltch, and Nilanjan Banerjee. 2025. Detection of cortical arousals in sleep using multimodal wearable sensors and machine learning. *Scientific Reports* 15, 1 (2025), 44089. doi:10.1038/s41598-025-27739-7

[21] Duri Lee, Heejeong Lim, Vedant Das Swain, and Uichin Lee. 2026. Understanding Behind the Smile of Emotion Workers: Detecting After-Call Stress in Call Agents. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*. 1–21.

[22] Hancong Liu, Sirish Shah, and Wei Jiang. 2004. On-line outlier detection and data cleaning. *Computers & chemical engineering* 28, 9 (2004), 1635–1647.

[23] Yunfei Luo, Yuliang Chen, Asif Salekin, and Tauhidur Rahman. 2025. Toward Foundation Model for Multivariate Wearable Sensing of Physiological Signals. *ACM Transactions on Computing for Healthcare* (2025). doi:10.1145/3803808

[24] Yee Jia Luwe, Chin Poo Lee, and Kian Ming Lim. 2022. Wearable sensor-based human activity recognition with hybrid deep learning model. In *Informatics*, Vol. 9. MDPI, 56.

[25] Mithun Saha, Maxwell A. Xu, Wanting Mao, Sameer Neupane, James M. Rehg, and Santosh Kumar. 2025. Pulse-PPG: An Open-Source Field-Trained PPG Foundation Model for Wearable Applications Across Lab and Field Settings. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* (2025). doi:10.1145/3749494

[26] Florian Schaule, Jan Ole Johanssen, Bernd Bruegge, and Vivian Loftness. 2018. Employing consumer wearables to detect office workers' cognitive load for interruption management. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 2, 1 (2018), 1–20.

[27] Marcin Strączkiewicz, Nancy W. Glynn, and Jaroslaw Harezlak. 2019. On placement, location and orientation of wrist-worn tri-axial accelerometers during free-living measurements. *Sensors* 19, 9 (2019), 2095.

[28] Hang Yuan, Shing Chan, Andrew P. Creagh, Catherine Tong, Aidan Acquah, David A. Clifton, and Aiden Doherty. 2024. Self-supervised learning for human activity recognition using 700,000 person-days of wearable data. *npj Digital Medicine* 7, 1 (2024), 91. doi:10.1038/s41746-024-01062-3