# Non-Contact Health Monitoring During Daily Personal Care Routines

Xulin Ma[1*], Jiankai Tang[2,3*], Zhang Jiang[2], Songqin Cheng[2], Yuanchun Shi[1,2], Dong Li[1], Xin Liu[4], Daniel McDuff[4], Xiaojing Liu[1†], Yuntao Wang[1,2,3†]

*Co-first Author, †Co-corresponding Author

1 Department of Computer Technology and Applications, Qinghai University, Qinghai, 810016. liuxj@qhu.edu.cn

2 National Key Laboratory of Human Factors Engineering, Beijing, 100094. yuntaowang@tsinghua.edu.cn

3 Department of Computer Science and Technology, Tsinghua University, Beijing, 100084. tjk24@mails.tsinghua.edu.cn

4 Paul G. Allen School of Computer Science & Engineering, University of Washington, Seattle WA, 98195, xliu0@cs.washington.edu

***Abstract*—Remote photoplethysmography (rPPG) enables non-contact, continuous monitoring of physiological signals and offers a practical alternative to traditional health sensing methods. Although rPPG is promising for daily health monitoring, its application in long-term personal care scenarios—such as mirror-facing routines in high-altitude environments—remains challenging due to ambient lighting variations, frequent occlusions from hand movements, and dynamic facial postures. To address these challenges, we present the Long-term Altitude Daily Health (LADH) dataset, the first long-term rPPG dataset containing 240 synchronized RGB and infrared (IR) facial videos from 21 participants across five common personal care scenarios, along with ground-truth PPG, respiration, and blood oxygen signals. Our experiments demonstrate that combining RGB and IR video inputs improves the accuracy and robustness of non-contact physiological monitoring, achieving a mean absolute error (MAE) of 4.99 BPM in heart rate estimation. Furthermore, we find that multi-task learning enhances performance across multiple physiological indicators simultaneously. Dataset and code are open at https://github.com/McJackTang/FusionVitals.**

***Index Terms*—non-contact, rPPG, health-monitoring**

## I. Introduction

Traditional methods for measuring vital signs, such as heart rate (HR), blood oxygen saturation ($SpO_2$), and respiratory rate (RR), often require bulky or intrusive equipment, which limits their practicality for continuous monitoring in daily life [1], [2]. Remote photoplethysmography (rPPG) provides a non-invasive alternative by using cameras to detect subtle changes in skin reflectance, enabling less obtrusive monitoring in everyday settings [3]. Despite these advantages, rPPG faces challenges in real-world environments, where physiological signals can be affected by factors such as clothing, hair, cosmetics, head movement, and changes in ambient lighting. This study focuses on addressing these challenges in daily personal care scenarios.

Although public benchmark datasets have contributed significantly to the advancement of rPPG research, most do not address the specific challenges found in daily personal care scenarios [4]. For example, commonly used rPPG datasets such as PURE [5], UBFC-rPPG [6], and SUMS [7] are mainly collected under resting-state conditions and may not capture the physiological variations that occur during routine personal care activities.

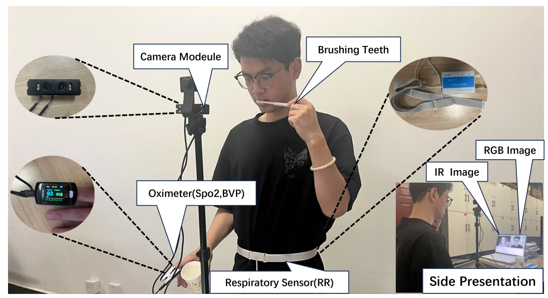


Fig. 1. The experimental setup of data collection while participants are brushing teeth.

To address these challenges, we present **L**ong-term **A**ltitude **D**aily **H**ealth (**LADH**) dataset, a multi-modal biosensing dataset for daily personal care activities. LADH contains 240 synchronized RGB and IR facial videos collected across five scenarios: sitting at rest, sitting during personal care (toothbrushing/hair-combing with facial occlusions), standing at rest, standing during personal care, and post-exercise. Eleven participants contributed data over a continuous **10-day** period, with all recordings synchronized at millisecond precision with ground-truth physiological signals from pulse oximeters and respiratory bands [7].

Our evaluation using both subject-independent and day-wise partitioning reveals that day-wise partitioning enhances measurement accuracy, while RGB and IR video fusion with multi-task learning significantly improves physiological signal prediction despite frequent facial occlusions.

Our main contributions include:

- The first long-term rPPG dataset featuring synchronized RGB and IR videos with ground-truth physiological measurements across daily personal care scenarios with realistic facial occlusions.
- Demonstration that multi-modal fusion of RGB and IR inputs improves non-contact physiological monitoring accuracy (achieving 4.99 BPM MAE) despite facial occlusions during toothbrushing and hair-combing.
- Evidence that multi-task learning enhances performance across multiple physiological indicators simultaneously in both heart and respiratory rate estimation.

TABLE I
DATASET COMPARISON

| Dataset | Videos | Camera-Position | Vitals | Long-term | Obscured |
|---|---|---|---|---|---|
| PURE [5] | 40 | Face | PPG/$SpO_2$ | ✗ | ✗ |
| UBFC-rPPG [6] | 42 | Face | PPG | ✗ | ✗ |
| MMPD [8] | 660 | Face | PPG | ✗ | ✗ |
| SUMS [7] | 80 | Face+Finger | PPG/$SpO_2$/RR | ✗ | ✗ |
| LADH | 240 | Face(RGB+IR) | PPG/$SpO_2$/RR | ✓ | ✓ |

## II. Related Works

### A. Physiological Sensing in Daily Personal Care Scenarios

With the development of non-contact physiological sensing technologies, daily personal care activities such as face washing, mirror viewing, and tooth brushing have become practical contexts for health monitoring due to their routine and accessible nature [9]. Regular monitoring of vital signs, including heart rate, blood pressure, and blood oxygen saturation, may support early detection and prevention of various health conditions [10]. Unlike controlled laboratory settings, these everyday activities provide opportunities for unobtrusive physiological data collection in real-world environments.

Previous studies have demonstrated the feasibility of smart mirror systems for non-contact, real-time heart rate measurement without external sensors [11], [12]. These results indicate that non-contact health monitoring systems can offer a passive, user-friendly, and privacy-preserving approach in daily personal care scenarios. However, there is limited research on the robustness of such systems in long-term, comprehensive user studies. The LADH dataset aims to address this gap by focusing on daily personal care routines in the real world.

### B. Physiological Sensing Datasets

Many widely used remote photoplethysmography (rPPG) datasets—such as PURE [5], UBFC-rPPG [6], MMPD [8], and SUMS [7]—have been collected under controlled laboratory conditions, typically involving static seated positions or limited facial movement. As a result, these datasets may not fully capture the dynamic variations present in daily personal care activities, such as tooth brushing or mirror-facing routines.

To address this limitation, we developed the LADH dataset to support the evaluation of rPPG methods in everyday personal care scenarios. LADH contains synchronized RGB and IR facial videos, along with ground-truth physiological signals, collected from 21 participants over 10 days in real-world environments. Compared to existing datasets, LADH introduces challenges such as facial occlusion and varying indoor lighting, providing a resource for studying the robustness of non-contact physiological signal measurement. Table I summarizes several related datasets for comparison.

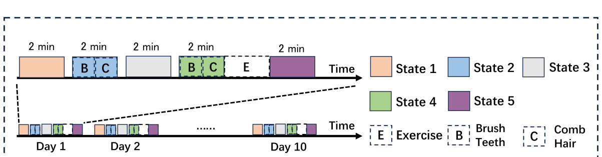


Fig. 2. A visual illustration of our daily data collection protocol. Participants have different activities across states.

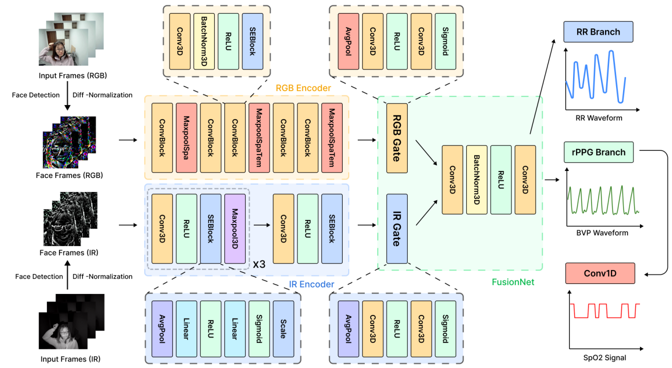


Fig. 3. FusionPhys Model with Input frames of facial RGB and facial IR. PPG, RR and SpO2 estimation tasks are trained simultaneously with a combined loss.

## III. Dataset

Approved by the Institutional Review Board, we recruited 21 graduate students (aged 23–28) as participants and conducted data collection in a simulated daily personal care scenario. The WN-12207K3321SM290 camera module was used to capture participants' facial videos in both RGB and IR modalities. As shown in Figure 1, physiological ground-truth signals were recorded using a CMS50E pulse oximeter for photoplethysmography (PPG) and $SpO_2$, and an HKH-11C respiratory sensor to monitor breathing patterns. Video recordings were acquired at a resolution of 640×480 pixels and a frame rate of 30 frames per second (FPS). PPG signals were sampled at 20 Hz, while RR signals were recorded at 50 Hz. All video frames and physiological signals were synchronously collected at millisecond-level precision using a data acquisition platform based on the PhysRecorder [13].

### A. Data Collection

The data collection protocol is illustrated in Figure 2. Data were collected from 21 participants in five scenarios, where 11 subjects participated in a 10-day longitudinal study, while the remaining 10 contributed single-day recordings. For the seated resting condition (state 1), participants wore an HKH-11C respiratory sensor on their abdomen and a CMS50E pulse oximeter on their left index finger while sitting upright facing the camera. Participants maintained minimal movement with a fixed gaze at the camera during the two-minute recording period.

In state 2, participants remained seated with the same equipment while performing toothbrushing and hair-combing activities for two minutes. State 3 involved a standing resting position with the same sensor configuration, also recorded for two minutes. In state 4, participants repeated the toothbrushing and hair-combing actions while standing for two minutes. After completing these four conditions, participants engaged in moderate physical exercise (squats, high knees, or breath-holding) to induce physiological changes. Immediately following exercise, participants were recorded for an additional two minutes in a seated position (state 5). The exercise component was included to produce variations in $SpO_2$ levels.

## IV. Method

### A. Input embedding

Building upon PhysNet [7], [14], our FusionPhysNet introduces an input embedding strategy that processes facial video from two modalities: RGB and IR. As shown in Figure 3, both video streams are pre-processed into frame sequences and passed through a shared 3D convolutional encoder for spatiotemporal feature extraction. By integrating features from both modalities, the model achieves a more comprehensive understanding of an individual's physiological state, effectively combining superficial and deeper physiological signals. This multimodal fusion framework significantly enhances the model's performance in complex real-world scenarios and establishes a more accurate and reliable foundation for vital sign estimation, including heart rate and blood oxygen saturation.

### B. Neural Network Model

The overall architecture of the FusionPhysNet is illustrated in Figure 3. We extend the PhysNet [14] backbone with a modality-aware fusion mechanism in FusionNet. Specifically, a gated feature selection strategy adaptively modulates the contribution of each modality based on global contextual representations. This design enables the model to dynamically emphasize the more informative modality under varying environmental conditions (e.g., changes in illumination), enhancing the robustness and generalizability of the physiological signal estimation framework.

### C. Joint-Training

Inspired by the multitask training framework [7], we incorporate RR estimation into the joint training mechanism, allowing simultaneous optimization of HR, $SpO_2$ and RR prediction tasks. Furthermore, we introduce a novel loss function specifically designed to jointly handle the objectives of HR, $SpO_2$ and RR estimation. By optimizing these three targets concurrently during training, the model is guided to learn a more comprehensive and informative feature representation, thus improving its capacity to capture complex physiological dynamics and improving the robustness of vital sign prediction. It is formulated as:

$$\begin{aligned}\text{Loss} &= \text{MSE}_{\text{BVP}} + \text{MSE}_{\text{RR}} \\ &\quad + 0.002 \times \text{MSE}_{\text{SpO}_2} \times (100 - \text{Mean}(\text{SpO}_2))\end{aligned}$$

where $\text{MSE}_{\text{BVP}}$ represents the mean squared error of the blood volume pulse (BVP) signal, $\text{MSE}_{\text{RR}}$ represents the mean squared error of the RR signal, and $\text{MSE}_{\text{SpO}_2}$ represents the mean squared error of the $SpO_2$ signal. This formulation aims to balance the precision of the BVP, RR and SpO2 measurements, leading to a more robust model for multimodal physiological signal measurement.

TABLE II
Intra-Dataset and Inter-Dataset Experiment

| Training Set | LADH | | SUMS | | PURE | |
|---|---|---|---|---|---|---|
| **Test Set** | MAE↓ | MAPE↓ | MAE↓ | MAPE↓ | MAE↓ | MAPE↓ |
| LADH | 8.15 | 9.19 | 16.93 | 18.20 | 17.00 | 18.78 |
| SUMS | 11.23 | 15.45 | 3.36 | 3.84 | 14.95 | 17.11 |
| PURE | 8.10 | 8.83 | 7.97 | 8.87 | 0.59 | 0.77 |

The table shows the cross-dataset experimental results of the LADH, SUMS [7], and PURE [5] datasets on the **PhysNet [14]** model.

TABLE III
Results of HR-$SpO_2$-RR Multi-Task Training by Subject

| Modality | HR Task | | SpO2 Task | | RR Task | |
|---|---|---|---|---|---|---|
| | MAE↓ | MAPE↓ | MAE↓ | MAPE↓ | MAE↓ | MAPE↓ |
| Both(Single Task) | 9.02 | 10.99 | <u>1.10</u> | <u>1.19</u> | 2.25 | 10.16 |
| RGB(Multi Task) | 9.34 | 12.08 | 1.29 | 1.39 | 3.08 | 13.78 |
| IR(Multi Task) | 12.99 | 15.73 | 1.23 | 1.33 | 2.41 | 11.20 |
| Both(Multi Task) | **<u>7.12</u>** | **<u>8.93</u>** | **1.14** | **1.23** | **<u>1.43</u>** | **<u>6.53</u>** |
| Gains | +1.90 | +2.06 | −0.04 | −0.04 | +0.82 | +3.63 |

MAE = Mean Absolute Error in HR estimation (Beats/Min), MAPE = Mean Percentage Error (%). <u>underline</u> means the best performance. **Gains** denote the improvements brought by multimodal and multitask learning.

## V. Results and Findings

All experiments were conducted using the improved training framework of rPPG-Toolbox [15]. The environment configuration included PyTorch 2.2.2+cuda12.1, Python 3.8, and NVIDIA A100 GPU. The experiments were performed with a learning rate of 9e-3, 30 epochs, and a batch size of 16.

We validated the effectiveness of LADH through both intra- and inter-dataset experiments using PhysNet [14]. As shown in Table II, LADH achieves reasonable intra-dataset accuracy (MAE=8.15 BPM) and maintains competitive cross-dataset performance when tested on SUMS and PURE, outperforming SUMS in cross-transfer to PURE (MAE 8.10 vs. 14.95). These results highlight LADH's dual role: it provides a strong baseline within its own domain while also exposing the inherent challenges of generalization across different datasets, underscoring its value as a benchmark for robust rPPG evaluation.

We then conducted two sets of experiments based on different dataset partitioning strategies within LADH: subject-wise partitioning and day-wise partitioning. We established different training tasks, including individual training for HR, $SpO_2$ and RR under multimodal input, joint training for HR-$SpO_2$-RR under multimodal input, joint training for HR-$SpO_2$-RR under RGB-only video input and IR-only video input.

**In the subject-wise partitioning experiment, multimodal fusion with joint training outperforms single-modality and single-task approaches, particularly for HR and RR estimation.** The dataset was partitioned such that data from 8 subjects were used for training, 3 subjects for validation, and an additional dataset from 10 individuals was reserved for testing. The results indicated significant improvements in the MAE for HR, which decreased from 9.02 to 7.12, reflecting a 21.06% error reduction, and for RR, which decreased from 2.25 to 1.43, reflecting a 36.44% error reduction. This suggests that multimodal fusion and joint training are more effective for periodic tasks like HR and RR, while $SpO_2$ does not exhibit clear periodic fluctuations and is inferred through indirect signals.

TABLE IV
RESULTS OF HR-$SpO_2$-RR MULTI-TASK TRAINING BY DAY

| Modality | HR Task | | SpO2 Task | | RR Task | |
|---|---|---|---|---|---|---|
| | MAE↓ | MAPE↓ | MAE↓ | MAPE↓ | MAE↓ | MAPE↓ |
| Both(Single Task) | 5.23 | 5.44 | 1.31 | 1.38 | 2.57 | 13.45 |
| RGB(Multi Task) | 5.73 | 5.77 | 1.35 | 1.43 | 1.99 | 9.12 |
| IR(Multi Task) | 8.35 | 8.98 | 1.28 | 1.36 | 1.51 | 6.74 |
| Both(Multi Task) | **4.99** | **5.21** | **1.29** | **1.37** | **2.24** | **11.38** |

MAE = Mean Absolute Error in HR estimation (Beats/Min), MAPE = Mean Percentage Error (%). underline means the best performance.

**In the day-wise partitioning experiment, multimodal fusion with joint training improves HR estimation, and multitask learning benefits $SpO_2$ and RR estimation.** In this experiment, data collected over 10 days were split into 7 days for training, 2 days for validation, and 1 day for testing. The results showed that for HR estimation, multimodal fusion with joint training outperformed single-modality and single-task approaches, reducing MAE from 5.23 to 4.99 (a 4.59% error reduction). In IR-based joint training, errors for $SpO_2$ and RR were reduced by 2.29% and 41.25%, respectively. This highlights the effectiveness of multimodal fusion for HR and multitask learning for $SpO_2$ and RR.

**Comparison of the subject-wise and day-wise experiments illustrates how day-wise analysis can improve the adaptability of models to individual user data.** While the subject-wise experiment shows strong performance for periodic tasks through multimodal fusion and joint training, the day-wise experiment emphasizes the ability of the model to adapt more closely to individual data. This could indicate that, in future personalized health monitoring systems, such as a health mirror, models can better accommodate daily variations and offer more tailored results to users, enhancing the accuracy of HR, RR, and $SpO_2$ estimation on an individual level.

## VI. LIMITATIONS

While this study provides valuable insights into the feasibility of non-contact health monitoring in daily personal care scenarios, it has certain limitations. The relatively small sample size and age range may not fully capture the diversity of physiological characteristics and behavioral patterns across the general population. Future research should aim to include a larger and more diverse participant pool and incorporate a broader range of real-world personal care behaviors to enhance the generalizability and robustness of the findings.

## VII. CONCLUSION

This study introduces the LADH dataset, the first long-term rPPG dataset with synchronized RGB, IR, and ground-truth physiological signals across realistic daily care scenarios. By releasing 240 multimodal videos with PPG, respiration, and $SpO_2$, LADH provides a valuable benchmark for robust non-contact health monitoring. We further propose FusionPhys, a modality-fusion and multi-task learning framework that improves accuracy and generalization, reducing HR and RR errors across both cross-subject and cross-day settings. FusionPhys reveals the complementary value of RGB and IR signals and the effectiveness of joint HR–$SpO_2$–RR prediction. This work lays a foundation for accurate, robust, and non-invasive health monitoring in everyday scenarios.

## ACKNOWLEDGMENT

This work is supported by the foundation of National Key Laboratory of Human Factors Engineering No. HFNKL2024W06, the National Natural Science Foundation of China No. 62366043 & 62472244, Beijing Natural Science Foundation No.QY24248, the National Key R&D Program of China No. 2024YFB4505500, No. 2024YFB4505503, the Tsinghua University Initiative Scientific Research Program No. 20257020004, and Qinghai University Research Ability Enhancement Project No. 2025KTSA05.

## REFERENCES

[1] Jiankai Tang, Kegang Wang, Hongming Hu, Xiyuxing Zhang, Peiyu Wang, Xin Liu, and Yuntao Wang. Alpha: Anomalous physiological health assessment using large language models. In *AI Health Summit*, 2023.

[2] Jiankai Tang, Kegang Wang, Yingke Ding, Jiatong Ji, Zeyu Wang, Xiyuxing Zhang, Ping Chen, Yuanchun Shi, and Yuntao Wang. A dataset and toolkit for multiparameter cardiovascular physiology sensing on rings. *arXiv e-prints*, pages arXiv–2505, 2025.

[3] Daniel McDuff. Camera measurement of physiological vital signs. *ACM Computing Surveys*, 55(9):1–40, 2023.

[4] Mingxuan Liu, Jiankai Tang, Haoxiang Li, Jiahao Qi, Siwei Li, Kegang Wang, Yuntao Wang, and Hong Chen. Spiking-physformer: Camera-based remote photoplethysmography with parallel spike-driven transformer. *Neural Networks*, 2024.

[5] Ronny Stricker, Steffen Müller, and Horst-Michael Gross. Non-contact video-based pulse rate measurement on a mobile service robot. In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, pages 1056–1062. IEEE, 2014.

[6] Serge Bobbia, Richard Macwan, Yannick Benezeth, Alamin Mansouri, and Julien Dubois. Unsupervised skin tissue segmentation for remote photoplethysmography. *Pattern Recognition Letters*, 124:82–90, 2019.

[7] Ke Liu, Jiankai Tang, Zhang Jiang, Yuntao Wang, Xiaojing Liu, Dong Li, and Yuanchun Shi. Summit vitals: Multi-camera and multi-signal biosensing at high altitudes. In *UIC*, 2024.

[8] Jiankai Tang, Kequan Chen, Yuntao Wang, Yuanchun Shi, Shwetak Patel, Daniel McDuff, and Xin Liu. Mmpd: Multi-domain mobile video physiology dataset. In *EMBC*, pages 1–5. IEEE, 2023.

[9] Jiankai Tang, Xinyi Li, Jiacheng Liu, Xiyuxing Zhang, Zeyu Wang, and Yuntao Wang. Camera-based remote physiology sensing for hundreds of subjects across skin tones. In *CHI'24 Workshop PhysioCHI'24*, 2024.

[10] Debjyoti Talukdar, Luis Felipe De Deus, and Nikhil Sehgal. Evaluation of a camera-based monitoring solution against regulated medical devices to measure heart rate, respiratory rate, oxygen saturation, and blood pressure. *Cureus*, 14(11), 2022.

[11] Ming-Zher Poh, Daniel McDuff, and Rosalind Picard. A medical mirror for non-contact health monitoring. In *ACM SIGGRAPH 2011 Emerging Technologies*, pages 1–1. 2011.

[12] Kolja Dobrovoljski, Mihaela Stoycheva, Vukan Turkulov, Shahina Begum, and Mobyen Uddin Ahmed. Smartmirror: An embedded non-contact system for health monitoring at home. In *Internet of Things Technologies for HealthCare*, volume 187, page 133. Springer, 2017.

[13] Kegang Wang, Yantao Wei, Jiankai Tang, Yuntao Wang, Mingwen Tong, Jie Gao, Yujian Ma, and Zhongjin Zhao. Camera-based hrv prediction for remote learning environments. In *UIC*, 2024.

[14] Zitong Yu, Xiaobai Li, and Guoying Zhao. Remote photoplethysmograph signal measurement from facial videos using spatio-temporal networks. In *Proc. BMVC*, 2019.

[15] Xin Liu, Girish Narayanswamy, Akshay Paruchuri, Xiaoyu Zhang, Jiankai Tang, Yuzhe Zhang, Roni Sengupta, Shwetak Patel, Yuntao Wang, and Daniel McDuff. rppg-toolbox: Deep remote ppg toolbox. In *Advances in Neural Information Processing Systems*, volume 36, 2024.