Rapid Adaptation of SpO_2 Estimation to Wearable Devices via Transfer Learning on Low-Sampling-Rate PPG

Zequan Liang[1], Ruoyu Zhang[2], Wei Shao[1], Krishna Karthik[2], Ehsan Kourkchi[2], Setareh Rafatirad[1], Houman Homayoun[2]

[1]Department of Computer Science, University of California, Davis, Davis, CA, U.S.A.
[2]Department of Electrical and Computer Engineering, University of California, Davis, Davis, CA, U.S.A.
Email: {zqliang, ryuzhang, wayshao, kbhaskar, ekay, srafatirad, hhomayoun}@ucdavis.edu

***Abstract*—Blood oxygen saturation (SpO_2) is a vital marker for healthcare monitoring. Traditional SpO_2 estimation methods often rely on complex clinical calibration, making them unsuitable for low-power, wearable applications. In this paper, we propose a transfer learning–based framework for the rapid adaptation of SpO_2 estimation to energy-efficient wearable devices using low-sampling-rate (25Hz) dual-channel photoplethysmography (PPG). We first pretrain a bidirectional Long Short-Term Memory (BiLSTM) model with self-attention on a public clinical dataset, then fine-tune it using data collected from our wearable We-Be band and an FDA-approved reference pulse oximeter. Experimental results show that our approach achieves a mean absolute error (MAE) of 2.967% on the public dataset and 2.624% on the private dataset, significantly outperforming traditional calibration and non-transferred machine learning baselines. Moreover, using 25Hz PPG reduces power consumption by 40% compared to 100Hz, excluding baseline draw. Our method also attains an MAE of 3.284% in instantaneous SpO_2 prediction, effectively capturing rapid fluctuations. These results demonstrate the rapid adaptation of accurate, low-power SpO_2 monitoring on wearable devices without the need for clinical calibration.**

***Index Terms*—SpO_2 estimation; photoplethysmography (PPG); wearable health monitoring; transfer learning; machine learning**

I. Introduction

Blood oxygen saturation (SpO_2) is a critical physiological parameter that reflects the percentage of oxygen-bound hemoglobin in the blood, providing essential insights into cardiovascular function [1], [2]. Pulse oximetry estimates SpO_2 non-invasively using photoplethysmography (PPG) sensors by analyzing the light absorption of pulsatile blood flow, typically red and infrared (IR) wavelengths [3].

In recent years, machine learning models have been employed for SpO_2 estimation from PPG signals. However, several challenges remain for practical deployment in wearable health monitoring systems. First, many existing approaches rely solely on single-channel input [4] and require high PPG sampling rates, which leads to increased power consumption. Second, traditional SpO_2 calibration on new devices often requires controlled clinical calibration, which is impractical for rapid deployment [5]. Lastly, spontaneous SpO_2 fluctuations are poorly represented in clinical datasets where saturation levels are generally maintained within stable ranges [5].

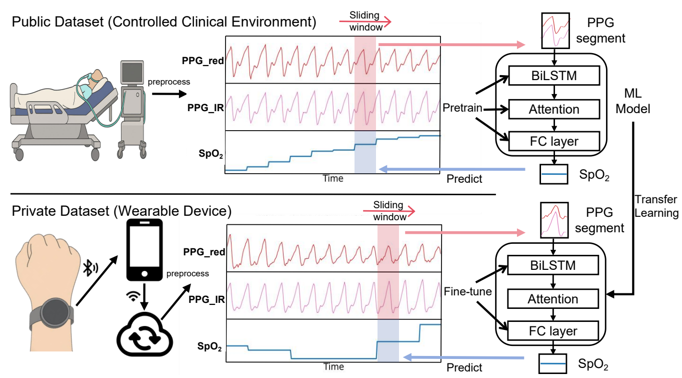

Fig. 1. SpO_2 estimation framework using transfer learning

To address these limitations, we propose a transfer learning–based framework for rapid adaptation of SpO_2 estimation to new wearable devices using low-rate dual-channel PPG. As shown in Fig. 1, our approach mainly pretrains a bidirectional Long Short-Term Memory (BiLSTM) model with self-attention [6] on a publicly available dataset collected in a controlled clinical environment, and then fine-tunes it using a private dataset collected from the wearable We-Be band [7].

This work makes the following key contributions: **(1) Low-Sampling-Rate PPG Enables Energy-Efficient Wearable Devices:** Our system operates at a sampling rate of 25Hz, reducing power consumption and extending the battery life of wearable devices for long-term health monitoring. **2) Transfer Learning Enables Rapid Deployment Across Devices:** We enable fast model adaptation on a wearable device through transfer learning by fine-tuning on small amounts of user-level data, eliminating the need for clinical calibration. **3) Capturing Instantaneous SpO_2 Fluctuations:** Our model is capable of detecting rapid SpO_2 changes, which are essential for early warning in conditions such as sleep apnea and respiratory distress.

II. Framework Overview

The overall framework is illustrated in Fig. 1. We utilize the OpenOximetry Dataset [5], a publicly available dataset collected under a medical-grade clinical environment. Red (660nm) and infrared (940nm) PPG signals were recorded at a sampling rate of 86Hz, along with reference SpO_2 values obtained by averaging readings from multiple pulse oximeters. To align with our We-Be band [8], which operates at 25Hz, we downsample the PPG signals from 86Hz to 25Hz by Fourier-domain resampling. After preprocessing, both red and IR signals are then segmented using a sliding window. Each resulting segment is paired with its corresponding SpO_2 value, enabling supervised learning. BiLSTM has been widely adopted for temporal feature extraction [9]. Our machine learning (ML) model includes a BiLSTM, a self-attention layer, and a fully connected (FC) layer. They are pretrained jointly to learn PPG patterns and regress the SpO_2 values.

To make the rapid adaptation, we collect a private dataset, as illustrated in Fig. 2, using the We-Be band equipped with PPG sensors (660nm red and 950nm IR, sampled at 25Hz), alongside an FDA-approved Masimo Rad-G pulse oximeter as the reference. To induce transient SpO_2 fluctuations, participants perform at least 3 cycles of breath-holding. We apply the same preprocessing and segmentation procedures to the private dataset and fine-tune the pretrained ML model. Specifically, we fine-tune the BiLSTM and the FC layer to regress the reference SpO_2 values better.

After transfer learning, the resulting model is deployed on the We-Be band for real-time SpO_2 estimation. It continuously processes dual-channel PPG segments and outputs a continuous stream of SpO_2 predictions.

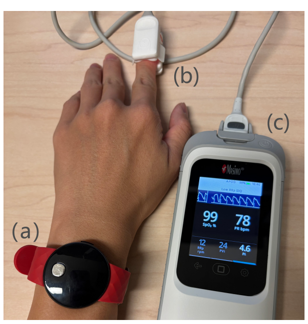

Fig. 2. Private wearable data collection. (a) We-Be band for PPG signals (b) Fingertip sensor attached to Masimo Rad-G (c) Masimo Rad-G for reference SpO_2

III. Data Preprocessing and Model Architecture

The data preprocessing pipeline is illustrated in Fig. 3(a). We begin by extracting the raw PPG signals from both red and infrared (IR) sensors. To remove baseline drift and high-frequency noise, we apply a band-pass filter with a frequency range of 0.5–12Hz. This filter separates the signal into its alternating current (AC) and direct current (DC) components: the DC component captures low-frequency baseline trends below 0.5Hz, while the AC component captures the pulsatile

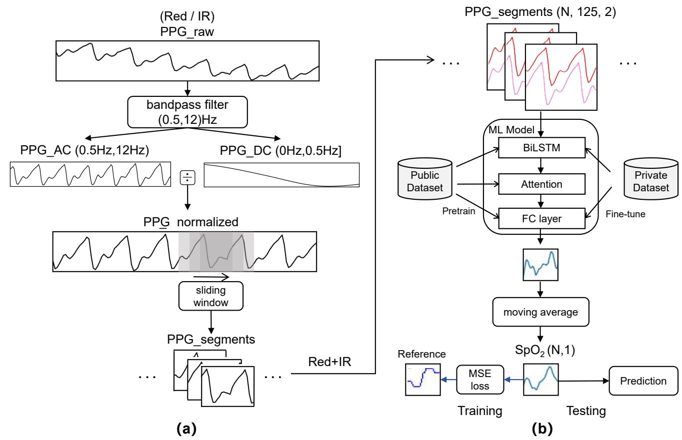

Fig. 3. Data processing (a) PPG preprocessing (b) Machine learning method

variation of blood flow within the 0.5–12Hz range. To normalize the PPG signal, we compute the ratio of the AC to DC components. Next, we segment the normalized signal using a sliding window of 5 seconds with a 1-second stride. Given the 25Hz sampling rate, each window contains 125 sample points.

As shown in Fig. 3(b), we concatenate the red and IR PPG segments along the last dimension, resulting in a segment of shape $(125, 2)$. Across all N segments, the input tensor to the model has shape $(N, 125, 2)$. Our ML model consists of three main components: a BiLSTM to learn temporal dependencies in the dual-channel PPG sequence, a self-attention layer that uses the internal BiLSTM features as the query, key, and value [6] to capture the global contextual relationship, and an FC layer to regress the final SpO_2 value. The model outputs a single SpO_2 prediction per window via the FC layer. A moving average filter with a window size of 5 is applied to the series of SpO_2 predictions, resulting in a final SpO_2 vector of shape $(N, 1)$. During training, we optimize the model using the mean squared error (MSE) loss between predicted and reference SpO_2 values. During testing, the output is used directly as the SpO_2 prediction.

We first pretrain all model layers on the public dataset collected under controlled clinical conditions. Then we use our private wearable dataset to fine-tune only the FC layer. Finally, we unfreeze the BiLSTM and continue training both the BiLSTM and the FC layer together. This two-stage fine-tuning strategy results in a robust and low-cost solution for SpO_2 estimation on wearable platforms.

IV. Experiment on Public Dataset

To pretrain the model for SpO_2 estimation, we first conducted experiments on the public OpenOximetry dataset. The dataset was split at the subject level into training and test sets using a 4:1 ratio. Additionally, 5-fold cross-validation was performed within the training set. To match the characteristics of our We-Be band, we downsampled the original PPG signals from 86 Hz to 25 Hz by Fourier-domain resampling after a 0.5-12Hz band-pass filter. All models were implemented in PyTorch and trained using the Adam optimizer with a

TABLE I
SpO$_2$ PREDICTION ERRORS ON PUBLIC OPENOXIMETRY DATASET

Method	25Hz		86Hz	
	MAE	*RMSE*	*MAE*	*RMSE*
Traditional	3.286	4.892	3.280	4.880
CNN_1d	5.899	7.047	5.687	6.997
Transformer	3.577	5.151	4.208	6.019
BiLSTM	3.216	4.787	3.049	4.551
BiLSTM+Attn	**2.967**	**4.393**	**2.910**	**4.331**

learning rate of 0.001. The batch size was set to 256, and each model was trained for 100 epochs using the MSE loss function. To address label imbalance in SpO$_2$, we also apply weighted sampling during fine-tuning. The SpO$_2$ labels were discretized into 10 bins, and sampling weights were assigned as the inverse bin frequencies.

We compared the following five methods:

- Traditional [10]: A traditional method that computes the R ratio for quadratic SpO$_2$ calibration:

$$R = \frac{AC_{\mathrm{RED}}/DC_{\mathrm{RED}}}{AC_{\mathrm{IR}}/DC_{\mathrm{IR}}} \tag{1}$$

- CNN_1d: A one-dimensional convolutional neural network that applies temporal convolution layers followed by a fully connected layer for regression.
- Transformer: A transformer encoder with 4 self-attention heads applied to segmented PPG for sequence modeling.
- BiLSTM: A two-layer bidirectional Long Short-Term Memory network that captures temporal dependencies in the PPG signal.
- BiLSTM+Attn: Based on BiLSTM, this model integrates a self-attention mechanism to emphasize internal features in the temporal representation.

Table I presents the performance of all methods on the public OpenOximetry dataset in terms of Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). Notably, the prediction error at 25Hz remains close to that before downsampling from 86Hz, indicating that key SpO$_2$ information is largely preserved. Even after downsampling to 25Hz, the BiLSTM+Attn model maintains the best overall performance, with an MAE of 2.967% and an RMSE of 4.393%. Fig. 4 illustrates a representative test case using 25Hz PPG. Compared to the reference SpO$_2$ (blue solid line), the traditional method (green dashed) fails to accurately track low SpO$_2$ levels below 85%. In contrast, the BiLSTM+Attn model (red dashed) closely follows the reference, demonstrating its effectiveness in capturing complex signal dynamics.

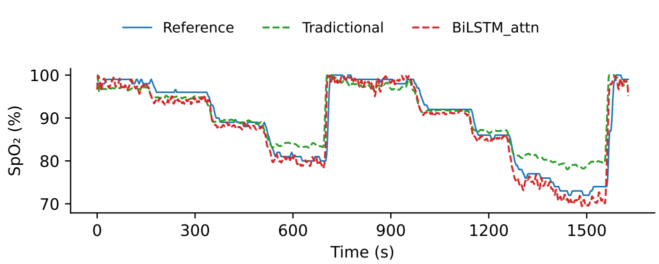

Fig. 4. SpO$_2$ prediction results on a representative test case from public OpenOximetry Dataset using 25Hz PPG

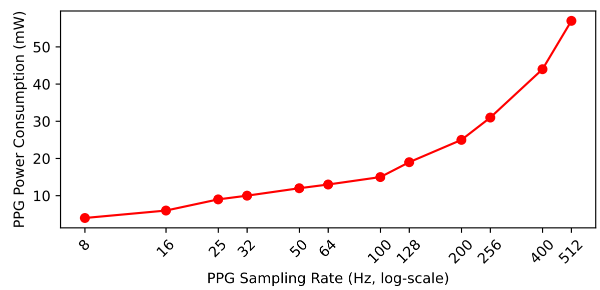

Fig. 5. Power consumption of the PPG sensor on the We-Be band

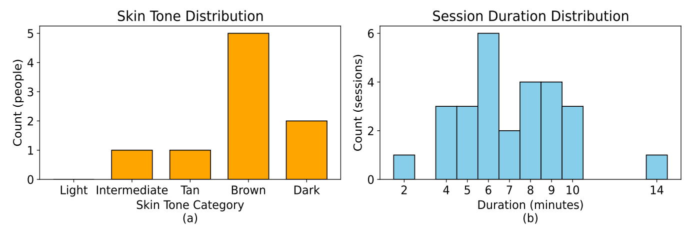

Fig. 6. Distributions of the private dataset (a) Skin tones (b) Session durations

V. EXPERIMENT ON PRIVATE DATASET

As shown in Fig. 5, we measured the power consumption of the PPG sensor across various sampling rates on the We-Be band. After subtracting the static power drawn by non-PPG components, the We-Be band consumes only around 9mW at 25Hz PPG, which is 40% lower than around 15mW required at the commonly used 100Hz PPG. This demonstrates the significant energy savings achieved by using a lower sampling frequency.

To evaluate the effectiveness of our model in non-clinical conditions, we conducted experiments on a private dataset collected using wearable devices. The dataset includes dual-channel 25Hz PPG signals from We-Be band, along with corresponding SpO$_2$ reference values from Masimo Rad-G. In total, we collected 27 sessions from 9 individuals, with 3 sessions per person. We also measured wrist skin color using a Vinckolor colorimeter to classify their skin tones [11]. Fig. 6 summarizes the distributions of skin tone categories and session durations. Due to the limited sample size, we employed a leave-one-subject-out (LOSO) cross-validation strategy to ensure subject-independent evaluation.

We compared the following methods as before:

- Traditional: The same calibration method as above.
- BiLSTM+Attn: The pretrained BiLSTM model with a self-attention layer, tested directly on the private dataset.
- BiLSTM+Attn+Transfer: A transfer learning approach where the pretrained BiLSTM with self-attention is fine-tuned on the private dataset for 150 epochs.

TABLE II
SpO$_2$ PREDICTION ERRORS ON PRIVATE WEARABLE DATASET USING 25HZ PPG

Method	MAE	RMSE	MAE$_{ins}$	RMSE$_{ins}$
Traditional	4.069	5.228	4.738	5.780
BiLSTM+Attn	2.717	3.717	4.635	5.734
BiLSTM+Attn+Transfer	**2.624**	**3.214**	**3.284**	**3.999**

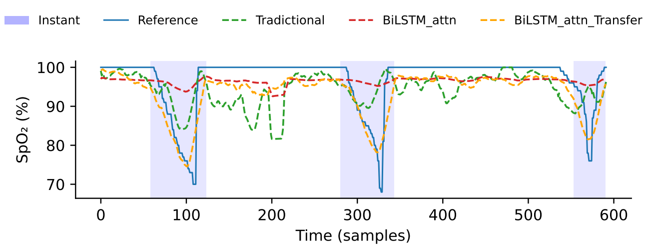

Fig. 7. SpO$_2$ prediction results on a representative test case from private wearable dataset

A significant oxygen desaturation event is typically defined as a $\geq 3\%$ drop in SpO$_2$ sustained for at least 10 seconds as measured by pulse oximetry [12]. To further assess performance during instantaneous SpO$_2$ fluctuations, we define instant zones as time intervals where the total variation (TV) in SpO$_2$ exceeds 3% within a sliding window of 10 seconds:

$$TV_i = \sum_{j=i}^{i+10} |y_{j+1} - y_j| \tag{2}$$

where y_j represents the reference SpO$_2$ value at time step j. A time window is flagged as an instant zone if $TV_i \geq 3\%$ at the starting index i. In these zones, we measured instantaneous errors MAE$_{ins}$ and RMSE$_{ins}$ to evaluate the model's ability to track sharp desaturation or resaturation events.

As shown in Table II, on the private dataset, both machine learning models outperformed the traditional calibration method. The BiLSTM+Attn+Transfer model achieved the best overall performance, with the lowest MAE of 2.624% and MAE$_{ins}$ of 3.284%. In addition, Fig. 7 illustrates a representative test case, highlighting the instant zones in purple. Compared to the reference SpO$_2$ (blue solid line), only the BiLSTM+Attn+Transfer model (orange dashed) closely follows the reference during rapid SpO$_2$ drops, effectively tracking transient desaturation..

VI. DISCUSSION

Our results demonstrate the accurate SpO$_2$ estimation using PPG signals sampled at 25Hz. Despite the reduced temporal resolution, the BiLSTM+Attn model outperforms traditional calibration methods. While pretraining on the public dataset offers a strong initialization, the model struggles to generalize to wearable data due to domain shift and hardware differences. Transfer learning significantly improves performance in these non-clinical settings. Moreover, when fine-tuned on small amounts of wearable data, BiLSTM-based models with attention effectively capture rapid desaturation events, making them valuable for applications such as sleep apnea detection and remote monitoring.

VII. CONCLUSION AND FUTURE WORK

We propose a transfer learning framework for rapid and accurate SpO$_2$ estimation on wearable devices using low-sampling-rate (25 Hz) PPG signals. By pretraining a BiLSTM with self-attention on a public clinical dataset and fine-tuning it on a small set of wearable data, our method eliminates the need for device-specific clinical calibration while substantially reducing power consumption. The proposed approach demonstrates strong performance on both public and private datasets, effectively capturing both average and instantaneous SpO$_2$ fluctuations and outperforming traditional calibration and non-transferred machine learning models. In future work, we will leverage temporal dependencies across sliding windows and employ better domain adaptation methods.

REFERENCES

[1] D. Evans, B. Hodgkinson, and J. Berry, "Vital signs in hospital patients: a systematic review," *International journal of nursing studies*, vol. 38, no. 6, pp. 643–650, 2001.

[2] R. Fang, R. Zhang, S. M. Hosseini, M. Faghih, S. Rafatirad, S. Rafatirad, and H. Homayoun, "Pain level modeling of intensive care unit patients with machine learning methods: An effective congeneric clustering-based approach," in *2022 4th International Conference on Intelligent Medicine and Image Processing*, pp. 89–95, 2022.

[3] J. Kumar V and K. A. Reddy, "Pulse oximetry for the measurement of oxygen saturation in arterial blood," *Studies in Skin Perfusion Dynamics: Photoplethysmography and Its Applications in Medical Diagnostics*, pp. 51–78, 2021.

[4] M. N. I. Shuzan, M. H. Chowdhury, M. E. Chowdhury, M. Murugappan, E. Hoque Bhuiyan, M. Arslane Ayari, and A. Khandakar, "Machine learning-based respiration rate and blood oxygen saturation estimation using photoplethysmogram signals," *Bioengineering*, vol. 10, no. 2, p. 167, 2023.

[5] N. Fong, M. S. Lipnick, E. Behnke, Y. Chou, S. Elmankabadi, L. Ortiz, C. S. Almond, I. Auchus, G. W. Burnett, R. Bisegerwa, *et al.*, "Open access dataset and common data model for pulse oximeter performance data," *Scientific data*, vol. 12, no. 1, pp. 1–8, 2025.

[6] W. Li, F. Qi, M. Tang, and Z. Yu, "Bidirectional lstm with self-attention mechanism and multi-channel features for sentiment classification," *Neurocomputing*, vol. 387, pp. 63–77, 2020.

[7] R. Zhang, R. Fang, M. Orooji, and H. Homayoun, "Introducing webe band: an end-to-end platform for continuous health monitoring," in *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1–5, IEEE, 2024.

[8] R. Fang, S. Hang, R. Zhang, C. Fang, S. Rafatirad, C. Hostinar, and H. Homayoun, "Validation of webe band during physical activities," in *2024 IEEE 20th International Conference on Body Sensor Networks (BSN)*, pp. 1–4, IEEE, 2024.

[9] Z. Liang, H. Luo, X. Chen, Z. Zhong, C. Fan, X. Song, B. Li, and J. Lv, "A multi-modal multi-expert framework for pain assessment in postoperative children," *IEEE Transactions on Affective Computing*, 2025.

[10] T. Guo, Z. Cao, Z. Zhang, D. Li, and M. Yu, "Reflective oxygen saturation monitoring at hypothenar and its validation by human hypoxia experiment," *Biomedical engineering online*, vol. 14, no. 1, p. 76, 2015.

[11] B. C. K. Ly, E. B. Dyer, J. L. Feig, A. L. Chien, and S. Del Bino, "Research techniques made simple: cutaneous colorimetry: a reliable technique for objective skin color measurement," *Journal of Investigative Dermatology*, vol. 140, no. 1, pp. 3–12, 2020.

[12] S. Kainulainen, B. Duce, H. Korkalainen, A. Oksenberg, A. Leino, E. S. Arnardottir, A. Kulkas, S. Myllymaa, J. Töyräs, and T. Leppänen, "Severe desaturations increase psychomotor vigilance task-based median reaction time and number of lapses in obstructive sleep apnoea patients," *European Respiratory Journal*, vol. 55, no. 4, 2020.