# SmartVitals-BP : Contactless Blood Pressure Estimation from Facial Videos via Multi-Scale Temporal Modeling and Metadata Fusion

Routhu Srinivasa Rao*, Aradhana Mishra, Anoushka Panigrahi, Sanjay Swain
CureBay, Bhubaneswar, Odisha, India, 751022
* Email: routhu.srinivasa@curebay.com

***Abstract*—Hypertension is a leading risk factor for cardiovascular diseases, yet continuous blood pressure (BP) monitoring remains inaccessible in many low-resource and rural settings due to device dependency, cost, and infrastructure constraints. This paper presents SmartVitals-BP, a deep learning-based, contactless framework for estimating systolic (SBP) and diastolic (DBP) blood pressure from standard RGB facial videos captured using commodity cameras. The proposed system leverages remote photoplethysmography (rPPG) along with patient demographic metadata to extract cardiovascular signals from subtle facial color variations. A multi-scale 1D Convolutional Neural Network (CNN) is employed to capture hierarchical temporal features, followed by a Bidirectional Long Short-Term Memory (BiLSTM) network with attention to model temporal dependencies and enhance discriminative signal patterns. To address the inherent imbalance in BP distributions, a distribution-aware composite loss that combines ordinal regression and variance regularization is introduced to preserve prediction fidelity across physiological ranges. Evaluated using a strict subject-independent 70:15:15 split, the model achieves a Mean Absolute Error (MAE) of 12.47 mmHg for SBP and 7.02 mmHg for DBP. Furthermore, on a challenging real-time clinical inference dataset comprising 46 unseen, in-the-wild samples, SmartVitals-BP demonstrates strong generalization, achieving an SBP MAE of 11.87 mmHg and a DBP MAE of 9.35 mmHg. These results highlight the potential of the framework for scalable, non-invasive, and device-independent BP screening, enabling population-scale preventive healthcare and digital health applications.**

## I. Introduction

Blood pressure (BP) is among the most critical physiological parameters for diagnosing and managing cardiovascular diseases (CVD), which remain the leading cause of mortality worldwide [1], [2]. Accurate, continuous monitoring of systolic and diastolic BP is essential for early hypertension detection and prevention of complications such as stroke and heart failure [3]. Conventional cuff-based methods are well established but inherently discontinuous, uncomfortable for extended use, and impractical for real-time monitoring [4]. Invasive intra-arterial techniques enable continuous measurement but are restricted to critical care environments [4].

Photoplethysmography (PPG) has emerged as a widely adopted modality for cuffless, non-invasive BP estimation. PPG sensors detect volumetric changes in blood flow by measuring light absorption variations at the skin surface, encoding morphological features related to arterial stiffness, vascular tone, and pulse wave velocity [5], [6]. Multi-sensor approaches combining PPG with electrocardiography (ECG) exploit pulse transit time (PTT) and pulse arrival time (PAT) as physiological proxies for BP [7]. The rapid advancement of deep learning has further transformed BP estimation, enabling end-to-end learning directly from raw waveforms without manual feature engineering, with architectures including CNNs, RNNs/LSTMs, residual networks, and graph neural networks all applied to this problem [8], [9].

In parallel, remote photoplethysmography (rPPG) extracts pulse waveforms from ordinary RGB camera video by detecting subtle skin color changes caused by pulsatile blood flow, opening the door to fully contactless BP estimation [10]. This is particularly attractive for telehealth, home monitoring, and large-scale screening. However, significant challenges remain: rPPG signals are considerably noisier than contact PPG due to illumination variation, head motion, and skin tone diversity [11], [12]; many published methods report misleadingly optimistic results by failing to perform subject-level train/test splits [13]; large databases such as MIMIC-II/III contain substantial proportions of corrupted signals requiring careful cleaning [14]; and camera-based sensing encodes biases against individuals with darker skin tones [15].

This paper presents SmartVitals-BP, an end-to-end deep learning framework for continuous, contactless blood pressure estimation from rPPG signals extracted from facial videos, combined with patient demographic metadata. The proposed architecture consists of a signal preprocessing pipeline, a multi-scale signal encoder with temporal modeling, and a metadata encoding branch, followed by a feature fusion and regression module. Separate models are trained for systolic (SBP) and diastolic (DBP) blood pressure using a distribution-aware composite loss to address physiological variability and data imbalance. The remainder of this paper is organized as follows: Section II reviews related work; Section III describes the proposed methodology; Section IV presents experimental results and ablation studies; and Section V concludes the paper.

## II. Related Work

Early cuffless BP estimation methods exploited the physiological relationship between arterial pressure and pulse wave propagation, linking PTT and PAT to BP via the Moens-Korteweg equation [16]. Gesche et al. applied PWV-derived

linear regression for sleep monitoring [17], and Kachuee et al. extended this with calibration-free algorithms on MIMIC-II [18], though both require multi-sensor setups and periodic recalibration. Feature-based methods followed, extracting morphological and spectral PPG properties for regression. Haddad et al. combined pulse width, amplitude, and heart rate with multi-linear regression for hypertension classification [19], while others trained DNNs on PPG derivatives [20]. Khalid et al. confirmed through systematic comparison that morphological features alone carry sufficient discriminative power for both SBP and DBP prediction [21].

End-to-end deep learning eliminated manual feature engineering by learning directly from raw waveforms. Slapnicar et al. proposed a spectrotemporal residual network with subject-specific calibration [22], Schlesinger et al. used a Siamese CNN on PPG spectrograms [23], and Ibtehaz and Rahman introduced PPG2ABP, a fully convolutional network translating PPG directly to arterial BP waveforms [24]. Recurrent architectures have been equally prominent: Eom et al. trained deep recurrent networks for long-term prediction [25], Li et al. proposed a real-time LSTM for continuous estimation [26], and Harfiya et al. demonstrated LSTM-based PPG-to-ABP translation [27]. GNNs have more recently been applied to physiological signals, with Zanchetta et al. representing multi-channel PPG/ECG windows as graphs and training TGAT models achieving MAEs of 6.34 and 1.87 mmHg for SBP and DBP respectively [28].

For remote and contactless estimation, rPPG extracts pulse signals from facial video via subtle skin color changes. Model-based methods such as CHROM and POS suppress lighting artifacts through physical reflection models [29], [30], while deep learning approaches provide superior robustness under real-world conditions [10]. Wuerich et al. found that POS and CHROM outperform standard RGB under illumination variation and motion, while all methods degraded for darker skin tones due to reduced photon reflectance [11]. Luo et al. achieved low mean BP error using transdermal imaging across 17 facial ROIs [31], and Jain et al. reported MAEs of approximately 3.9 and 3.7 mmHg for SBP and DBP using PCA-derived rPPG with linear regression [32].

Data quality and evaluation methodology are critical but frequently overlooked. Schrumpf et al. revealed that models trained on MIMIC-III concentrate accuracy at frequent BP bins while incurring large errors at hypo- and hypertensive extremes, underscoring the need for subject-level train/test splits to avoid data leakage [13]. Salah et al. introduced a two-level PPG-ABP cleaning pipeline combining spectral alignment, physiological thresholding, and PCA-based outlier detection, demonstrating that carefully cleaned smaller datasets outperform larger uncleaned ones [14]. Liu et al. highlighted that most public datasets exclude irregular physiological states such as atrial fibrillation, limiting clinical generalizability, and that camera-based methods encode systematic biases against darker skin tones [15]. Algorithmic remedies including few-shot personalization [33] and synthetic data augmentation [34] have been proposed to address these disparities.

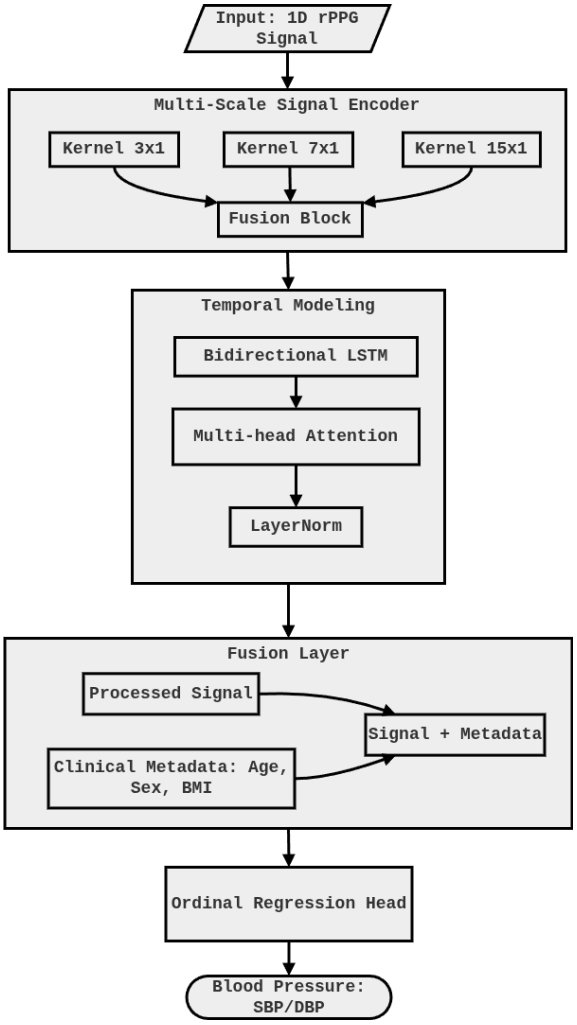


Fig. 1. Architecture of the proposed SmartVitals-BP framework

## III. Proposed Work

The proposed SmartVitals-BP framework is an end-to-end deep learning architecture for contactless blood pressure estimation using remote photoplethysmography (rPPG) signals and subject-specific metadata. As illustrated in Fig. 1, the framework integrates signal-driven representations with demographic priors into a unified prediction model. The architecture comprises three main components: a multi-scale signal encoder for hierarchical temporal feature extraction, a temporal modeling module based on BiLSTM with attention to capture sequential dependencies, and a multimodal feature fusion network that combines signal and metadata representations for final blood pressure estimation.

### *A. Video Preprocessing and rPPG Extraction*

Facial videos are processed using MediaPipe FaceMesh to detect dense facial landmarks for each frame. Based on these landmarks, three anatomically consistent regions of interest (ROIs)—forehead, left cheek, and right cheek—are defined using fixed landmark indices.

For each ROI, a polygonal mask is constructed, and the spatial average of RGB pixel intensities is computed, resulting in three temporal RGB signals. To extract the pulsatile component, the Chrominance (CHROM) method is applied independently to each ROI:

$$CHROM = 1.5 \cdot R_{norm} - G_{norm} - 0.5 \cdot B_{norm} \quad (1)$$

The resulting signals are fused using:

$$s(t) = 0.5 \cdot s_{\text{forehead}}(t) + 0.25 \cdot s_{\text{left}}(t) + 0.25 \cdot s_{\text{right}}(t) \quad (2)$$

The fused rPPG signal is resampled to a fixed length of 1000 samples and normalized to zero mean and unit variance.

### B. Multi-Scale Signal Encoder

To capture diverse temporal characteristics, the rPPG signal is processed through three parallel 1D convolutional branches with kernel sizes of 3, 7, and 15. This multi-scale design enables simultaneous modeling of fine-grained pulse morphology and broader temporal dynamics.

Each branch consists of convolutional layers with batch normalization and ReLU activation, followed by max pooling and dropout. The outputs are concatenated to form a unified multi-scale representation.

### C. Temporal Modeling

The multi-scale features are passed through a two-layer bidirectional LSTM (hidden size 128) to model temporal dependencies. A multi-head self-attention mechanism is subsequently applied to refine temporal representations by emphasizing informative segments. The final temporal embedding is obtained via mean pooling.

### D. Feature Fusion and Prediction

The outputs of the temporal encoder and subject-specific metadata (age, sex, BMI) are concatenated and passed through a fully connected fusion layer to learn cross-modal interactions in a shared latent space.

Blood pressure estimation is formulated using an ordinal regression framework, where the continuous BP range is discretized into a set of ordered bins. The model predicts cumulative probabilities for these bins using sigmoid-activated logits, and the final continuous BP estimate is obtained by summing the bin-wise probabilities scaled by the bin width.

### E. Training Objective

The model is optimized using a composite loss function:

$$\mathcal{L} = \alpha \mathcal{L}_{\text{Ordinal}} + \beta \mathcal{L}_{\text{Huber}} + \gamma \mathcal{L}_{\text{CCC}} + \epsilon \mathcal{L}_{\text{VarReg}} \quad (3)$$

This composite objective is designed to address the inherent imbalance in blood pressure distributions by enforcing ordinal consistency while preserving output variability through regularization. Specifically, the ordinal loss captures the ordered nature of BP values, the Huber loss provides robustness to outliers, the Concordance Correlation Coefficient (CCC) loss improves agreement between predicted and ground-truth distributions, and the variance regularization term prevents prediction collapse by maintaining physiological variability.

Separate models are trained for SBP and DBP using the AdamW optimizer with cosine annealing learning rate scheduling.

## IV. Experimentation and Results

### A. Datasets

We evaluate the proposed framework on two data distributions:

1) **MCD-rPPG Dataset :** The primary experiments were conducted on the recently released MCD-rPPG dataset introduced by Egorov et al [35]. The dataset contains 3,600 synchronized recordings collected from 600 subjects under both resting and post-exercise conditions using multiple consumer-grade cameras. Each recording includes synchronized facial videos, 100 Hz PPG signals, arterial blood pressure measurements, ECG, oxygen saturation ($SpO_2$), respiratory rate, body temperature, and additional physiological biomarkers. To ensure a fair evaluation and eliminate identity leakage, a strict subject-independent split of 70% for training, 15% for validation, and 15% for testing was adopted.
2) **Real-Time Clinical Dataset:** To further assess generalization in practical settings, an independent real-time clinical dataset comprising 46 samples was collected under unconstrained conditions. The recordings include variations in illumination, subject motion, facial pose, and camera quality, providing a challenging benchmark for evaluating the robustness of the proposed framework.

### B. Evaluation Metrics

Performance is evaluated using Mean Absolute Error (MAE) and Standard Deviation Ratio (StdR), defined as:

$$\text{Std Ratio} = \frac{\sigma_{\text{pred}}}{\sigma_{\text{gt}}} \quad (4)$$

While MAE measures prediction accuracy, StdR evaluates the ability of the model to preserve physiological variability in blood pressure estimation.

### C. Results and Analysis

*1) Overall Performance:* We first evaluate the overall performance of the proposed framework on both controlled and real-world settings. The results are summarized in Table I.

TABLE I
Mean Absolute Error (MAE) on Test Distributions

| Dataset | SBP MAE | DBP MAE |
|---|---|---|
| MCD-rPPG Test Split | 12.47 | 7.02 |
| Real-Time Inference ($N = 46$) | 11.87 | 9.35 |

On the subject-independent test split of the MCD-rPPG dataset, the model achieves MAEs of 12.47 mmHg for SBP and 7.02 mmHg for DBP, demonstrating competitive performance for contactless blood pressure estimation. When

evaluated on the real-time clinical dataset, the model maintains strong generalization, achieving an SBP MAE of 11.87 mmHg and a DBP MAE of 9.35 mmHg. The improvement in SBP suggests that the model effectively captures dominant hemodynamic trends, while the relatively higher DBP error reflects increased sensitivity to environmental noise and physiological variability.

*2) Impact of Metadata:* We analyze the contribution of subject-specific metadata (age, sex, BMI) by comparing the signal-only model with the metadata-enabled variant. The results are presented in Table II.

TABLE II
COMPARISON OF SIGNAL-ONLY AND METADATA-FUSED ARCHITECTURES

| Model Configuration | SBP MAE [mmHg] | DBP MAE [mmHg] | Metadata |
|---|---|---|---|
| Without Metadata | 13.58 | 7.90 | No |
| **With Metadata** | **12.47** | **7.02** | **Yes** |

The inclusion of metadata leads to a consistent improvement in performance, reducing MAE for both SBP and DBP. This indicates that demographic attributes provide complementary physiological context, enabling the model to better distinguish similar rPPG patterns across individuals.

*3) Architectural Ablation Study:* To identify the contribution of individual components, we perform a component-wise ablation study on the proposed architecture. Each variant is trained for 80 epochs, and performance is evaluated using both MAE and StdR. The results are summarized in Table III.

TABLE III
ARCHITECTURAL ABLATION STUDY

| Case | Removed Component | SBP MAE ↓ | SBP StdR ↑ | DBP MAE ↓ | DBP StdR ↑ |
|---|---|---|---|---|---|
| **Ref** | All components enabled | **12.47** | **0.64** | **7.02** | **0.58** |
| A1 | Multi-Scale Encoder | 12.53 | 0.60 | 7.30 | 0.56 |
| A2 | Temporal Attention | 12.49 | 0.63 | 7.05 | 0.57 |
| A3 | BiLSTM | 12.96 | 0.62 | 7.86 | 0.60 |
| A4 | Ordinal Regression | 12.41 | 0.49 | 6.84 | 0.43 |
| A5 | Variance Regularization | 12.57 | 0.61 | 7.20 | 0.54 |

The results reveal that the BiLSTM module plays a critical role in capturing temporal dependencies, as its removal leads to the largest degradation in performance. The multi-scale encoder and attention mechanisms contribute to improved feature representation, though their impact is comparatively moderate.

Notably, removing ordinal regression significantly reduces the StdR, indicating a collapse in prediction variability despite marginal improvements in MAE. This highlights the importance of distribution-aware learning in preserving physiologically meaningful predictions. Similarly, variance regularization contributes to maintaining output diversity, preventing regression-to-the-mean behavior.

*4) Comparison with Existing Methods:* To assess the effectiveness of the proposed framework, we compare SmartVitals-BP with representative blood pressure estimation approaches evaluated on the same MCD-rPPG benchmark. For fair comparison, the baseline methods were implemented using the same subject-independent train-validation-test split adopted in this work. The results are summarized in Table IV. The

TABLE IV
COMPARISON WITH THE EXISTING WORKS

| **Method** | **SBP MAE** | **DBP MAE** |
|---|---|---|
| Schrumpf et al [13] | 14.16 | 9.31 |
| Egorov et al [35] | 13.92 | 8.19 |
| **SmartVitals-BP (Proposed)** | **12.47** | **7.02** |

proposed SmartVitals-BP framework achieves the lowest error among all evaluated methods, improving SBP estimation by 1.45 mmHg and DBP estimation by 1.17 mmHg compared with the strongest baseline. These improvements demonstrate the benefit of combining multi-scale temporal modeling, metadata fusion, and distribution-aware optimization for contactless blood pressure estimation from facial videos.

### *D. Discussion*

The experimental results provide several key insights. First, incorporating demographic metadata significantly improves prediction stability and accuracy by introducing essential physiological context. Second, temporal modeling through BiLSTM is crucial for capturing sequential dynamics in rPPG signals, while convolutional and attention modules enhance feature extraction and refinement. Finally, distribution-aware objectives such as ordinal regression and variance regularization are essential for maintaining physiological variability, ensuring that the model produces clinically meaningful predictions rather than collapsing to average estimates.

Another notable observation is the stronger generalization performance for systolic blood pressure compared with diastolic blood pressure on the real-time dataset. This may be attributed to the stronger relationship between systolic pressure and pulse waveform morphology, while diastolic pressure is often more sensitive to peripheral vascular resistance and environmental noise. Future work will investigate larger and more diverse populations, advanced transformer-based temporal modeling, and multi-task learning strategies that jointly estimate blood pressure and additional physiological biomarkers available within the MCD-rPPG dataset

## V. CONCLUSION

In this paper, we presented SmartVitals-BP, a contactless deep learning framework for estimating blood pressure from facial videos. The proposed approach demonstrates that combining multi-scale temporal modeling with demographic priors enables reliable extraction of cardiovascular information from noisy rPPG signals. A key finding of this work is that distribution-aware learning plays a critical role in maintaining physiological fidelity, as reflected by improved variance preservation alongside competitive error metrics. The integration of metadata further stabilizes predictions by providing essential clinical context, particularly in scenarios

with similar signal patterns across individuals. Experimental results across both controlled and real-world datasets highlight the robustness and generalization capability of the proposed framework, supporting its potential for scalable, non-invasive blood pressure screening. Overall, this study indicates that contactless BP estimation can move closer to practical deployment when model design accounts for both signal complexity and physiological variability.

## REFERENCES


[1] M. Amini, F. Zayeri, and M. Salehi, "Trend analysis of cardiovascular disease mortality, incidence, and mortality-to-incidence ratio: results from global burden of disease study 2017," *BMC public health*, vol. 21, no. 1, p. 401, 2021.

[2] S. S. Lim, T. Vos, A. D. Flaxman, G. Danaei, K. Shibuya, H. Adair-Rohani, M. A. AlMazroa, M. Amann, H. R. Anderson, K. G. Andrews *et al.*, "A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the global burden of disease study 2010," *The lancet*, vol. 380, no. 9859, pp. 2224–2260, 2012.

[3] Y.-T. Zhang, Y.-L. Zheng, W.-H. Lin, H.-Y. Zhang, and X.-L. Zhou, "Challenges and opportunities in cardiovascular health informatics," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 3, pp. 633–642, 2013.

[4] R. Mukherjee, S. Ghosh, B. Gupta, and T. Chakravarty, "A literature review on current and proposed technologies of noninvasive blood pressure measurement," *Telemedicine and e-Health*, vol. 24, no. 3, pp. 185–193, 2018.

[5] G. Martínez, N. Howard, D. Abbott, K. Lim, R. Ward, and M. Elgendi, "Can photoplethysmography replace arterial blood pressure in the assessment of blood pressure?" *Journal of clinical medicine*, vol. 7, no. 10, p. 316, 2018.

[6] M. Elgendi, R. Fletcher, Y. Liang, N. Howard, N. H. Lovell, D. Abbott, K. Lim, and R. Ward, "The use of photoplethysmography for assessing hypertension," *NPJ digital medicine*, vol. 2, no. 1, p. 60, 2019.

[7] M. S. Tanveer and M. K. Hasan, "Cuffless blood pressure estimation from electrocardiogram and photoplethysmogram using waveform based ann-lstm network," *Biomedical Signal Processing and Control*, vol. 51, pp. 382–392, 2019.

[8] Y. Kurylyak, F. Lamonaca, and D. Grimaldi, "A neural network-based method for continuous blood pressure estimation from a ppg signal," in *2013 IEEE International instrumentation and measurement technology conference (I2MTC)*. IEEE, 2013, pp. 280–283.

[9] J. Esmaelpoor, M. H. Moradi, and A. Kadkhodamohammadi, "A multistage deep neural network model for blood pressure estimation using photoplethysmogram signals," *Computers in Biology and Medicine*, vol. 120, p. 103719, 2020.

[10] W. Chen and D. McDuff, "Deepphys: Video-based physiological measurement using convolutional attention networks," in *Proceedings of the european conference on computer vision (ECCV)*, 2018, pp. 349–365.

[11] C. Wuerich, K. Heinrich, C. Wiede, and K. Seidl, "Comparative analysis of color models for improved rppg signals in remote blood pressure measurement," in *Current Directions in Biomedical Engineering*, vol. 10, no. 4. De Gruyter, 2024, pp. 690–693.

[12] A. V. Moço, S. Stuijk, and G. De Haan, "New insights into the origin of remote ppg signals in visible light and infrared," *Scientific reports*, vol. 8, no. 1, p. 8501, 2018.

[13] F. Schrumpf, P. Frenzel, C. Aust, G. Osterhoff, and M. Fuchs, "Assessment of deep learning based blood pressure prediction from ppg and rppg signals," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 3820–3830.

[14] M. Salah, O. A. Omer, L. Hassan, M. Ragab, A. M. Hassan, and A. Abdelreheem, "Beat-based ppg-abp cleaning technique for blood pressure estimation," *IEEE Access*, vol. 10, pp. 55 616–55 626, 2022.

[15] X. Liu, S. Patel, and D. McDuff, "Rgb camera-based physiological sensing: Challenges and future directions," *arXiv preprint arXiv:2110.13362*, 2021.

[16] C. F. Wippermann, D. Schranz, and R. G. Huth, "Evaluation of the pulse wave arrival time as a marker for blood pressure changes in critically ill infants and children," *Journal of Clinical Monitoring*, vol. 11, no. 5, pp. 324–328, 1995.

[17] H. Gesche, D. Grosskurth, G. Küchler, and A. Patzak, "Continuous blood pressure measurement by using the pulse transit time: comparison to a cuff-based method," *European journal of applied physiology*, vol. 112, no. 1, pp. 309–315, 2012.

[18] M. Kachuee, M. M. Kiani, H. Mohammadzade, and M. Shabany, "Cuffless blood pressure estimation algorithms for continuous health-care monitoring," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 4, pp. 859–869, 2016.

[19] S. Haddad, A. Boukhayma, and A. Caizzone, "Continuous ppg-based blood pressure monitoring using multi-linear regression," *IEEE journal of biomedical and health informatics*, vol. 26, no. 5, pp. 2096–2105, 2021.

[20] Y.-C. Hsu, Y.-H. Li, C.-C. Chang, and L. N. Harfiya, "Generalized deep neural network model for cuffless blood pressure estimation with photoplethysmogram signal only," *Sensors*, vol. 20, no. 19, p. 5668, 2020.

[21] S. G. Khalid, J. Zhang, F. Chen, and D. Zheng, "Blood pressure estimation using photoplethysmography only: comparison between different machine learning approaches," *Journal of healthcare engineering*, vol. 2018, no. 1, p. 1548647, 2018.

[22] G. Slapničar, N. Mlakar, and M. Luštrek, "Blood pressure estimation from photoplethysmogram using a spectro-temporal deep neural network," *Sensors*, vol. 19, no. 15, p. 3420, 2019.

[23] O. Schlesinger, N. Vigderhouse, D. Eytan, and Y. Moshe, "Blood pressure estimation from ppg signals using convolutional neural networks and siamese network," in *ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2020, pp. 1135–1139.

[24] N. Ibtehaz, S. Mahmud, M. E. Chowdhury, A. Khandakar, M. Salman Khan, M. A. Ayari, A. M. Tahir, and M. S. Rahman, "Ppg2abp: Translating photoplethysmogram (ppg) signals to arterial blood pressure (abp) waveforms," *Bioengineering*, vol. 9, no. 11, p. 692, 2022.

[25] H. Eom, D. Lee, S. Han, Y. S. Hariyani, Y. Lim, I. Sohn, K. Park, and C. Park, "End-to-end deep learning architecture for continuous blood pressure estimation using attention mechanism," *Sensors*, vol. 20, no. 8, p. 2338, 2020.

[26] Y.-H. Li, L. N. Harfiya, K. Purwandari, and Y.-D. Lin, "Real-time cuffless continuous blood pressure estimation using deep learning model," *Sensors*, vol. 20, no. 19, p. 5606, 2020.

[27] L. N. Harfiya, C.-C. Chang, and Y.-H. Li, "Continuous blood pressure estimation using exclusively photopletysmography by lstm-based signal-to-signal translation," *Sensors*, vol. 21, no. 9, p. 2952, 2021.

[28] F. Zanchetta, A. Simonetti, G. Faglioni, A. Malagoli, R. Fioresi *et al.*, "A geometric deep learning approach to blood pressure regression." in *Geomedia Workshop 2022 in Proceedings of Machine Learning Research*, 2022, pp. N–A.

[29] G. De Haan and V. Jeanne, "Robust pulse rate from chrominance-based rppg," *IEEE transactions on biomedical engineering*, vol. 60, no. 10, pp. 2878–2886, 2013.

[30] W. Wang, A. C. Den Brinker, S. Stuijk, and G. De Haan, "Algorithmic principles of remote ppg," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 7, pp. 1479–1491, 2016.

[31] H. Luo, D. Yang, A. Barszczyk, N. Vempala, J. Wei, S. J. Wu, P. P. Zheng, G. Fu, K. Lee, and Z.-P. Feng, "Smartphone-based blood pressure measurement using transdermal optical imaging technology," *Circulation: Cardiovascular Imaging*, vol. 12, no. 8, p. e008857, 2019.

[32] M. Jain, S. Deb, and A. V. Subramanyam, "Face video based touchless blood pressure and heart rate estimation," in *2016 IEEE 18th international workshop on multimedia signal processing (MMSP)*. IEEE, 2016, pp. 1–5.

[33] X. Liu, Z. Jiang, J. Fromm, X. Xu, S. Patel, and D. McDuff, "Metaphys: few-shot adaptation for non-contact physiological measurement," in *Proceedings of the conference on health, inference, and learning*, 2021, pp. 154–163.

[34] D. McDuff, J. Hernandez, E. Wood, X. Liu, and T. Baltrusaitis, "Advancing non-contact vital sign measurement using synthetic avatars," *arXiv preprint arXiv:2010.12949*, 2020.

[35] K. Egorov, S. Botman, P. Blinov, G. Zubkova, A. Ivaschenko, A. Kolsanov, and A. Savchenko, "Gaze into the heart: A multi-view video dataset for rppg and health biomarkers estimation," in *Proceedings of the 33rd ACM International Conference on Multimedia*, 2025, pp. 13 053–13 059.