Hybrid modeling of photoplethysmography for non-invasive monitoring of cardiovascular parameters

Emanuele Palumbo* EMANUELE.PALUMBO@INF.ETHZ.CH
ETH Zurich

Sorawit Saengkyongam S_SAENGKYONGAM@APPLE.COM
Maria R. Cervera M_CERVERA@APPLE.COM
Jens Behrmann J_BEHRMANN@APPLE.COM
Andrew C. Miller ACMILLER@APPLE.COM
Guillermo Sapiro GSAPIRO@APPLE.COM
Christina Heinze-Deml C_HEINZEDEML@APPLE.COM
Antoine Wehenkel AWEHENKEL@APPLE.COM
Apple

Abstract

Continuous cardiovascular monitoring can play a key role in precision health. However, some fundamental cardiac biomarkers of interest, including stroke volume and cardiac output, require invasive measurements, e.g., arterial pressure waveforms (APW). As a non-invasive alternative, photoplethysmography (PPG) measurements are routinely collected in hospital settings. Unfortunately, the prediction of key cardiac biomarkers from PPG instead of APW remains an open challenge, further complicated by the scarcity of annotated PPG measurements. As a solution, we propose a hybrid approach that uses hemodynamic simulations and unlabeled clinical data to estimate cardiovascular biomarkers directly from PPG signals. Our hybrid model combines a conditional variational autoencoder trained on paired PPG-APW data with a conditional density estimator of cardiac biomarkers trained on labeled simulated APW segments. As a key result, our experiments demonstrate that the proposed approach can detect fluctuations of cardiac output and stroke volume and outperform a supervised baseline in monitoring temporal changes in these biomarkers.

Keywords: Cardiovascular Monitoring, Hemodynamics, Hybrid Learning, Generative Learning, Simulation-based Inference

Data and Code Availability In this work we make use of in-silico (simulated) and in-vivo (real-world) data. The in-vivo data is drawn from publicly available datasets (UCIBP (Kachuee et al., 2015) and VitalDB (Lee et al., 2022)) of hospitalized patients, while the in-silico dataset includes hemodynamic waveforms (Manduchi et al., 2024) generated using OpenBF, an open-source 1D blood flow solver (Benemerito et al., 2024).

Institutional Review Board (IRB) This work does not require IRB approval.

1. Introduction

Continuous monitoring of cardiovascular parameters such as cardiac output (CO) and stroke volume (SV) has proven essential to evaluate function, guide treatment, and detect hemodynamic instability in critical care (Pinsky, 2007; Nguyen and Squara, 2017). However, current clinical assessment relies on invasive techniques that entail risk and require medical supervision (Joosten et al., 2017; Arya et al., 2022; Evans et al., 2009). Photoplethysmography (PPG), which captures cardiovascular information non-invasively and is increasingly accessible through consumer wearables (Charlton et al., 2022; Weng et al., 2024), offers a promising alternative that could enable safer monitoring in critical care and potentially extend capabilities beyond hospital settings (Orini et al., 2023; Williams et al., 2023). Yet, monitoring cardiac biomarkers such as SV and CO from PPGs remains challenging due to the scarcity of labeled data and the indirect, not fully understood relationship between PPG signals and these parameters.

* Work done during an internship at Apple.

© 2025

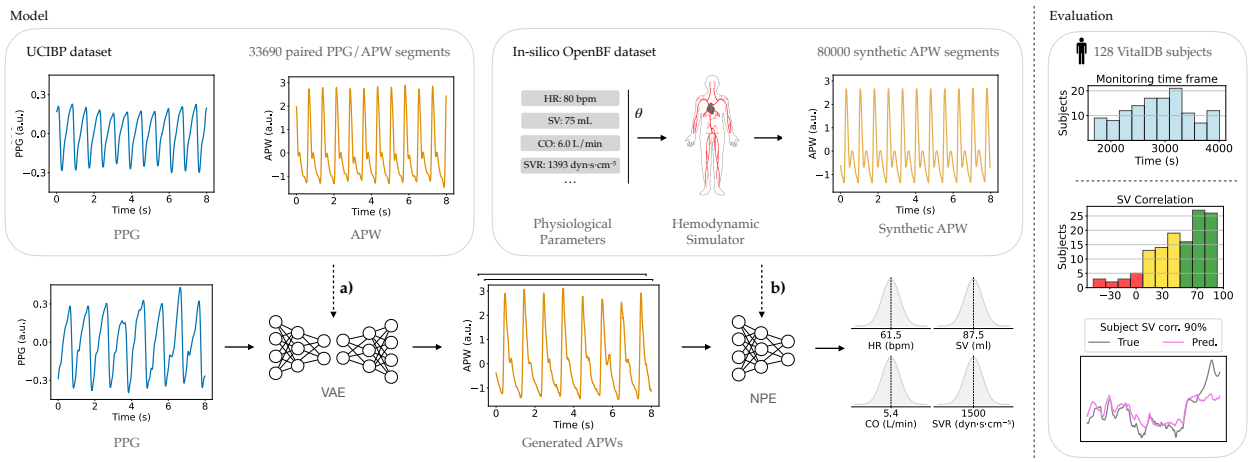


Figure 1: Proposed hybrid framework for non-invasive monitoring of cardiovascular parameters from PPG signals. The framework combines simulated and real-world data across two key components: **a)** a conditional VAE, trained on real-world paired APW and PPG segments, producing plausible APW signals from PPG inputs; **b)** a density estimator trained on a in-silico dataset of hemodynamic simulations inferring cardiovascular biomarkers from generated APW signals. The framework is validated in-vivo on unseen VitalDB data, using per-subject Spearman correlation between ground truth and predicted values of cardiovascular biomarkers.

In the context of cardiovascular monitoring, physiological simulations (Charlton et al., 2019; Melis et al., 2017) emerge as a potential strategy to mitigate the scarcity of in-vivo (i.e. real-patient) labeled data with cardiovascular biomarker annotations, using in-silico (i.e. simulated) data. In particular, recent work demonstrated the value of hemodynamic simulators to obtain arterial pressure waveform (APW) signals for inferring cardiovascular biomarkers in in-vivo data (Manduchi et al., 2024). While closely related to PPG signals, APWs are obtained invasively and thus provide a more direct view of central hemodynamics. By contrast, PPG reflects peripheral vascular changes through a complex relationship influenced by temperature, sweat, skin properties, and motion artifacts (Meier et al., 2024). Windkessel models attempt to describe the APW–PPG relationship, but they are too constrained to capture the full complexity of real-world data. Consequently, Manduchi et al. (2024) reported mixed results when applying the same strategy to PPG, despite its effectiveness on APWs.

To address these limitations, we propose a novel hybrid learning strategy that combines hemodynamic simulations with in-vivo data to estimate cardiac biomarkers directly from PPG signals. As in Manduchi et al. (2024), we first use a simulator of APWs to train a neural posterior estimator (NPE, Lueckmann et al., 2017) to model the relationship between APWs and cardiovascular parameters. To learn the transfer function between APW and PPG modalities, we then train a generative model on a real-world dataset of paired PPG-APW signals. Taken together, our contributions are: **(i)** a novel hybrid approach leveraging both in-silico simulations and in-vivo biosignals for PPG-based cardiovascular assessment; **(ii)** validation on unseen in-vivo data for the monitoring of stroke volume and cardiac output changes; and **(iii)** a demonstration that physics-based simulation enables robust cardiovascular parameter estimation from PPG signals.

2. Method

Our goal is to predict some cardiovascular parameters of interest given a finger PPG measurement, for which we propose a novel approach illustrated in Figure 1. We formalize this task as learning a conditional density estimator of $p(\theta \mid \mathbf{y})$, where $\mathbf{y} \in \mathbb{R}^T$ denotes a T-timestep PPG measurement and $\theta \in \mathbb{R}^d$ the parameters of interest. To solve this task, we are given **1.** a large dataset of labeled simulated APWs $\mathcal{D}_s := \{\theta_i, \mathbf{x}_i\}_{i=1}^{N_s}$, where $\theta_i \sim \pi(\theta)$ are sampled from a prior distribution π, and $\mathbf{x}_i \sim \tilde{p}(\mathbf{x} \mid \theta = \theta_i)$ are generated by running the simulator on θ_i (where $\tilde{p}$ emphasizes that the simulator approximates the real world); and

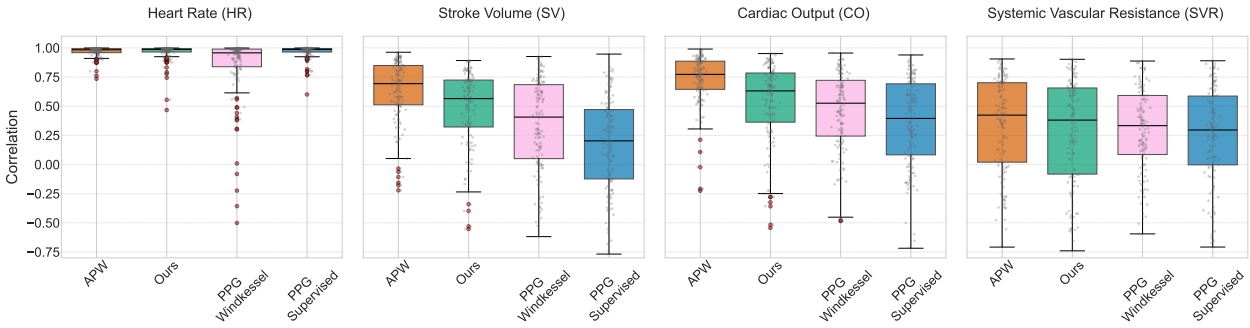


Figure 2: Box plots illustrating the per-subject Spearman correlation coefficients for the estimation of four key cardiovascular parameters. Our proposed hybrid approach (green) is compared against three relevant baselines: (i) *APW* (orange), which uses the ground-truth invasive APWs for inference with estimator trained in-silico; (ii) *PPG Windkessel* (pink), an estimator trained on labeled in-silico PPGs obtained with a Windkessel model approximation (Manduchi et al., 2024); and (iii) *PPG Supervised* (blue), an estimator trained directly on a subset of labeled VitalDB PPG data. Average results across independent runs with standard deviations are reported in Table 2.

2. a dataset $\mathcal{D}_r = \{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^{N}$ of simultaneous real-world APW and PPG measurements. In addition, for evaluation, we use a labeled dataset $\mathcal{D}_e := \{\theta_i, \mathbf{y}_i\}_{i=1}^{N_e}$ containing PPGs paired with ground truth parameter values.

Estimating $p(\theta \mid \mathbf{y})$. We approximate $p(\theta \mid \mathbf{y})$ by ignoring direct dependencies between parameters θ and PPG measurements $\mathbf{y}$, leading to

$$p(\theta \mid \mathbf{y}) \approx \mathbb{E}_{p(\mathbf{x}|\mathbf{y})}[p(\theta \mid \mathbf{x})].$$

We then learn two estimators, $p_{\phi_s}(\theta \mid \mathbf{x})$ and $p_{\phi_r}(\mathbf{x} \mid \mathbf{y})$ on the labeled simulation dataset $\mathcal{D}_s$ and paired APW-PPG dataset $\mathcal{D}_r$, respectively.

Deep generative modeling of $p(\mathbf{x} \mid \mathbf{y})$ from $\mathcal{D}_r$. We train a conditional variational auto-encoder (VAE, Kingma and Welling, 2014) on $\mathcal{D}_r$ to estimate $p(\mathbf{x} \mid \mathbf{y})$, the mapping from PPG to APW segments. In particular, we assume a latent generative model $p_{\phi_r}(\mathbf{x} \mid \mathbf{y}) := \mathbb{E}_{p(\mathbf{z})}[p_{\phi_r}(\mathbf{x} \mid \mathbf{z}, \mathbf{y})]$, where the latent $\mathbf{z} \sim p(\mathbf{z}) := \mathcal{N}(0, I)$ aims to capture unobserved physiological variations impacting the relationship between APWs and PPGs. In our experiments, we use 1D-CNN architectures and Gaussian densities to parameterize the encoder $q_{\phi_r}(\mathbf{z} \mid \mathbf{x}, \mathbf{y})$ and decoder $p_{\phi_r}(\mathbf{x} \mid \mathbf{z}, \mathbf{y})$ networks. We train these two networks by optimizing the Evidence Lower Bound objective on the the UCI Cuff-Less Blood Pressure Estimation (UCIBP) dataset (Kachuee et al., 2015), containing 333,690 paired APW and PPG 8-second segments from ICU patients. We provide details on the data format, architecture and training settings in appendix A.2.1.

Neural posterior estimation of $p(\theta \mid \mathbf{x})$. Following Manduchi et al. (2024), we train a neural posterior estimator (NPE) $p_{\phi_s}(\theta \mid \mathbf{x})$ by optimizing directly $\sum_{i=1}^{N_s} \log p_{\phi_s}(\theta_i \mid \mathbf{x}_i)$ via stochastic gradient ascent. To make $p_{\phi_s}(\theta_i \mid \mathbf{x}_i)$ differentiable we parameterize it with a conditional normalizing flow (NF, Rezende and Mohamed, 2015) composed of 3 steps of autoregressive and affine transformations and a convolutional neural network that encodes the APW. For our experiments, we reproduce the dataset from Manduchi et al. (2024), which contains $\approx$ 32,000 synthetic APW segments paired with parameter values. Appendix A.2.2 further documents how we train the NPE model.

Forming predictions with $p_\phi(\theta \mid \mathbf{y})$. For a given PPG segment $\mathbf{y}$, we first sample M plausible corresponding APW segments $\mathbf{x}_m \sim p_{\phi_r}(\mathbf{x} \mid \mathbf{y}, \mathbf{z}_m)$, where $\mathbf{z}_m \sim \pi$. Then, from each corresponding estimated posterior distribution $p_{\phi_s}(\theta \mid \mathbf{x}_m)$ we draw K samples. To obtain a point estimate we average the obtained $M \times K$ samples, and their standard deviation is used as a measure of uncertainty.

3. Results

We evaluate our approach on an entirely separate dataset, namely the VitalDB dataset (Lee et al., 2022), comprising APW and PPG signals from 128 patients undergoing non-cardiac surgery, labeled with cardiovascular biomarkers. Our evaluation focuses on four key cardiovascular parameters: Heart Rate (HR), Stroke Volume (SV), Cardiac Output (CO),

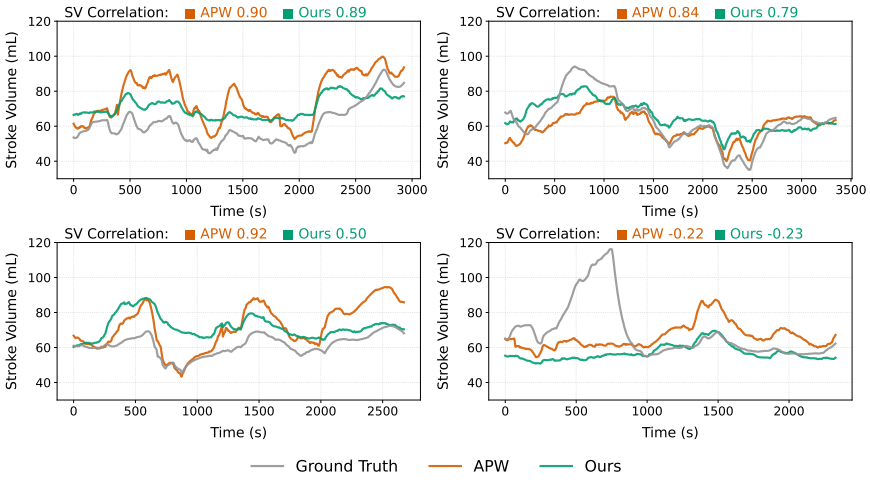


Figure 3: SV ground truth (gray), predicted values with our approach (green), and redictions obtained with invasive inference from ground truth APW (orange), on different subjects from the VitalDB dataset.

and Systemic Vascular Resistance (SVR). As a quantitative metric we report the per-patient Spearman correlation between ground truth and predicted values for a given cardiovascular parameter over the time frame that a patient is monitored.

Usefulness of hybrid modeling. Figure 2 compares our hybrid approach with three relevant baselines, described in the caption. First, all methods perform well at predicting HR. This confirms that our pipeline is capable, in principle, of predicting cardiovascular parameters whose effect is faithfully captured by the simulator. More interestingly, our method achieves a median per-subject correlation above 0.5 for both SV and CO, significantly higher than what Manduchi et al. (2024) obtained using a prescribed Windkessel transfer function between APWs and PPGs. Directly using groundtruth APWs as input to the predictive model, as in the APW baseline, expectedly outperforms our method, which instead relies on APW estimates. However, our method outperforms a PPG-supervised baseline in this task despite not having seen any in-vivo data with cardiovascular parameter labels, in contrast to this baseline. This comparison underscores the value of hemodynamic simulations for estimation of cardiovascular biomarkers from PPG inputs, and the effectiveness of combining prescribed simulations with learned components. Unlike purely model-based approaches, which often suffer from oversimplified assumptions (such as the Windkessel model), our hybrid modeling strategy has the potential to mitigate the absence of large labeled datasets.

Usefulness of generative modeling. We demonstrate that learning the PPG-to-APW transfer function with a conditional generative model offers benefits over deterministic models for the downstream tasks we consider in the experiments. Estimating APW from PPG is inherently difficult because PPG signals reflect peripheral vascular changes and are significantly influenced by noise, motion, and other confounding factors, making their relationship to invasive APWs complex and indirect. Hence deterministic models must compromise on this ambiguity, leading to a suboptimal information bottleneck for the parameters of interest. Table 1 in Appendix B.1 compares our method with a deterministic model and shows that deterministic models consistently achieve suboptimal performance. Furthermore, Figure 4 shows that prediction uncertainty tracks signal quality, highlighting the value of generative models for uncertainty modeling.

On the difficulties of absolute value prediction. In Figure 3, we examine the SV predictions for a subset of patients. While our approach reliably captures temporal trends in SV (e.g., it captures SV and CO peaks accurately), predicting the absolute values of such complex hemodynamic parameters remains a persistent challenge (Manduchi et al., 2024). Our results in Table 3 confirm this observation, showing that our method, as the other compared methods in this work, achieves limited ef-

fectiveness for absolute value prediction. This challenge stems from a combination of factors: (i) our current estimator performs a per-segment prediction and lacks a notion subject, thus overlooking unobservable, subject-specific physiological characteristics that influence baseline SV; and (ii) despite its sophistication, the underlying hemodynamic simulator may have inherent model misspecification when applied to real-world data. As future work, predictive model calibration on a small set of labeled data shall improve absolute value prediction. Furthermore, enabling personalized predictions, e.g., by introducing a notion of subject in the simulations and via the PPG-to-APW model, is another promising avenue.

4. Conclusion

In this work we use a hybrid modeling approach to infer cardiovascular parameters from in-vivo PPG signals. Compared to purely data-driven approaches that struggle due to limited labeled data, our method achieves promising results by incorporating simulations and sidestepping the need for invasive and costly annotations. While other existing hybrid approaches for cardiovascular modeling either embed physical properties as structural constraints within neural networks (Sel et al., 2023; Zhang et al., 2025) or augment traditional physiological models with data-driven components (Nazaret et al., 2023), our method incorporates physical knowledge in the model through SBI. While SBI relies on accurate forward models, in practice these may not be available to fully characterize each aspect of a complex process. We address this challenge and integrate a data-driven generative model which approximates the PPG-to-APW mapping, for which an accurate physics-based characterization is currently lacking. Our results contribute to characterizing the informativeness of PPG signals for predicting cardiac biomarkers, and could extend beyond the ones considered in our experiments. While our results are promising in monitoring temporal trends, absolute value prediction of complex biomarkers remains challenging, and is a key direction for future work. Future work may also explore alternative generative approaches for the PPG-to-APW mapping, or investigate different architectural choices. Finally, a similar learning strategy than the one used here for finger PPG could extend to other modalities, including wearable PPG, and open the door to passive and long-term cardiac biomarker monitoring.

References

Virendra K Arya, Waiwel Al-Moustadi, and Vikas Dutta. Cardiac output monitoring - invasive and noninvasive. *Current Opinion in Critical Care*, 2022.

I Benemerito, A Melis, A Wehenkel, and A Marzo. openbf: an open-source finite volume 1d blood flow solver. *Physiological Measurement*, 2024.

Peter H. Charlton, Jorge Mariscal Harana, Samuel Vennin, Ye Li, Phil Chowienczyk, and Jordi Alastruey. Modeling arterial pulse waves in healthy aging: a database for in silico evaluation of hemodynamics and pulse wave indexes. *American Journal of Physiology-Heart and Circulatory Physiology*, 2019.

Peter H. Charlton, Panicos A. Kyriacou, Jonathan Mant, Vaidotas Marozas, Phil Chowienczyk, and Jordi Alastruey. Wearable photoplethysmography for cardiovascular monitoring. *Proceedings of the IEEE*, 2022.

D. C. Evans, V. A. Doraiswamy, M. P. Prosciak, M. Silviera, M. J. Seamon, V. Rodriguez Funes, J. Cipolla, C. F. Wang, S. Kavuturu, D. A. Torigian, C. H. Cook, D. E. Lindsey, S. M. Steinberg, and S. P. Stawicki. Complications associated with pulmonary artery catheters: A comprehensive clinical review. *Scandinavian Journal of Surgery*, 2009.

Adham Hendy and Şerban Bubunek. Pulse waveform hemodynamic monitoring devices: recent advances and the place in goal-directed therapy in cardiac surgical patients. *Romanian Journal of Anaesthesia and Intensive Care*, 2016.

A. Joosten, O. Desebbe, K. Suehiro, L. S.-L. Murphy, M. Essiet, B. Alexander, M.-O. Fischer, L. Barvais, L. Van Obbergh, D. Maucort-Boulch, and M. Cannesson. Accuracy and precision of non-invasive cardiac output monitoring devices in perioperative medicine: a systematic review and meta-analysis. *BJA: British Journal of Anaesthesia*, 2017.

Mohamad Kachuee, Mohammad Mahdi Kiani, Hoda Mohammadzade, and Mahdi Shabany. Cuff-less high-accuracy calibration-free blood pressure estimation using pulse transit time. In *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2015.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *International Conference on Learning Representations*, 2014.

Hyung-Chul Lee, Yoonsang Park, Soo Bin Yoon, Seong-Mi Yang, Dongnyeok Park, and Chul-Woo Jung. Vitaldb, a high-fidelity multi-parameter vital signs database in surgical patients. *Scientific Data*, 2022.

Jan-Matthis Lueckmann, Pedro J. Goncalves, Giacomo Bassetto, Kaan Öcal, Marcel Nonnenmacher, and Jakob H. Macke. Flexible statistical inference for mechanistic models of neural dynamics. In *Advances in Neural Information Processing Systems*, 2017.

Laura Manduchi, Antoine Wehenkel, Jens Behrmann, Luca Pegolotti, Andy C. Miller, Marco Cuturi, Ozan Sener, Guillermo Sapiro, and Jörn-Henrik Jacobsen. Leveraging Cardiovascular Simulations for In-Vivo Prediction of Cardiac Biomarkers. In *ArXiv prepring:arXiv:2412.1754*, 2024.

Gerard R. Manecke. Edwards flotrac™ sensor and vigileo™ monitor: easy, accurate, reliable cardiac output assessment using the arterial pulse wave. *Expert Review of Medical Devices*, 2005.

Manuel Meier, Berken Utku Demirel, and Christian Holz. WildPPG: A real-world PPG dataset of long continuous recordings. In *The Thirty-eight Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2024.

Alessandro Melis, Richard H. Clayton, and Alberto Marzo. Bayesian sensitivity analysis of a 1d vascular model with gaussian process emulators. *International Journal for Numerical Methods in Biomedical Engineering*, 2017.

Achille Nazaret, Sana Tonekaboni, Gregory Darnell, Shirley You Ren, Guillermo Sapiro, and Andrew C. Miller. Modeling personalized heart rate response to exercise and environmental factors with wearables data. *npj Digital Medicine*, 2023.

Lee S. Nguyen and Pierre Squara. Non-invasive monitoring of cardiac output in critical care medicine. *Frontiers in Medicine*, 2017.

Michele Orini, Stefan van Duijvenboden, William J Young, Julia Ramírez, Aled R Jones, Andrew Tinker, Patricia B Munroe, and Pier D Lambiase. Premature atrial and ventricular contractions detected on wearable-format electrocardiograms and prediction of cardiovascular events. *European Heart Journal - Digital Health*, 2023.

Michael R. Pinsky. Hemodynamic evaluation and monitoring in the icu. *CHEST*, 2007.

Danilo Jimenez Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *International Conference on Machine Learning*, 2015.

Mohammed Saeed, Mauricio Villarroel, Andrew T Reisner, Gari D Clifford, Li-Wei H Lehman, George Moody, Thomas Heldt, Theerawit H Kyaw, Benjamin Moody, and Roger G Mark. Multiparameter intelligent monitoring in intensive care ii (mimic-ii): A public-access intensive care unit database. *Critical care medicine*, 39(5):952–960, 2011. doi: 10.1097/CCM.0b013e31820a92c6.

Kaan Sel, Amirmohammad Mohammadi, Roderic I. Pettigrew, and Roozbeh Jafari. Physics-informed neural networks for modeling physiological time series for cuffless blood pressure estimation. *npj Digital Medicine*, 2023.

Wei-Hung Weng, Sebastien Baur, Mayank Daswani, Christina Chen, Lauren Harrell, Sujay Kakarmath, Mariam Jabara, Babak Behsaz, Cory Y. McLean, Yossi Matias, Greg S. Corrado, Shravya Shetty, Shruthi Prabhakara, Yun Liu, Goodarz Danaei, and Diego Ardila. Predicting cardiovascular disease risk using photoplethysmography and deep learning. *PLOS Global Public Health*, 2024.

Gareth J Williams, Abdulaziz Al-Baraikan, Frank E Rademakers, Fabio Ciravegna, Frans N van de Vosse, Allan Lawrie, Alexander Rothman, Euan A Ashley, Martin R Wilkins, Patricia V Lawford, Stig W Omholt, Ulrik Wisløff, D Rodney Hose, Timothy J A Chico, Julian P Gunn, and Paul D Morris. Wearable technology and the cardiovascular system: the future of patient assessment. *The Lancet Digital Health*, 2023.

Yaowen Zhang, Libera Fresiello, Peter H. Veltink, Dirk W. Donker, and Ying Wang. A physiological-model-based neural network framework for blood pressure estimation from photoplethysmography signals. In *ArXiv preprint:arXiv:2502.03731*, 2025.

Appendix A. Experimental details

A.1. Datasets details

In this section we provide additional details for the in-silico and in-vivo datasets used in this work. It is worth noting that all PPG datasets considered in this work consist of finger PPG signals.

In-silico datasets We utilize an in-silico dataset of simulated Arterial Pressure Waveforms (APWs) generated using the OpenBF simulator, an open-source 1D blood flow solver (Benemerito et al., 2024), recently introduced by Manduchi et al. (2024). This dataset consists of 8-second 125Hz-frequency APW segments, synthesized from a comprehensive set of physiological parameters, that characterize the cardiovascular state for 80,000 virtual subjects. These parameters characterize heart function (e.g. Heart Rate, Stroke Volume, Peak Flow Time), arterial properties (e.g. arteries diameter/length) and vascular beds. The simulation process incorporates a stochastic noise model, combining Gaussian and red noise, to mimic the noise present in real measurements. Finally, the generated signals undergo a band-pass filtering. For the comparisons in Section 3 we use a PPG in-silico dataset, also introduced in Manduchi et al. (2024). Similar to APW generation, cardiovascular states for 80,000 patients are characterized by a comprehensive set of physiological parameters, and the PPG model is modeled by a simplistic Windkessel model (Manduchi et al., 2024; Charlton et al., 2019). While 80,000 APW and PPG signals were generated for both in-silico datasets, samples with anomalous systolic and diastolic bloop pressure were filtered out, leaving $\approx 32,000$ waveforms after this filtering.

In-vivo datasets We use two different in-vivo datasets in this work. The UCI Cuff-Less Blood Pressure Estimation (UCIBP) dataset (Kachuee et al., 2015) provides paired APW and PPG 8-second segments from ICU patients in the MIMIC-II dataset (Saeed et al., 2011), without subject information. The signals are sampled at 125Hz frequency. Note that the biosignals in this dataset are not annotated with any cardiovascular labels. For our evaluations in Section 3 we use the VitalDB dataset, which includes intraoperative APW and PPG signals from patients undergoing non-cardiac surgery in the Seoul National University Hospital with associated cardiovascular parameters. As cardiovascular biomarker labels we consider Heart Rate (HR), Stroke Volume (SV), Cardiac Output (CO), and Systemic Vascular Resistance (SVR). As in Manduchi et al. (2024), we process the VitalDB data to remove faulty signals (e.g. containg NaNs or implausible minimum or maximum values). We split the resulting APW and PPG waveforms in 8-second segments to match the lenght of simulated/UCIBP data.

A.2. Model and Implementation Details

A.2.1. Conditional VAE model

We use 1D-CNNs to parameterize the encoder and decoder of our conditional VAE model. We split the UCI Cuff-Less Blood Pressure Estimation (UCIBP) dataset as follows. We use $333,690$ paired APW-PPG segments from this dataset, randomly split in training (80%) and validation (20%) sets. We set the latent space size of the VAE to 128 dimensions. The VAE model is trained using the Adam Optimizer by maximizing the ELBO objective referenced in Section 2 for 300 epochs, with a learning rate of 0.0005, batch size of 256, and early stopping on the validation dataset.

A.2.2. NPE model

To invert the forward process of the OpenBF APW simulator, mapping $\theta \in \mathbb{R}^d$ to a given APW segment $\mathbf{x} \in \mathbb{R}^T$, we follow recent work (Manduchi et al., 2024) and train a Neural Posterior Estimator (NPE) to approximate the posterior distribution $p(\theta \mid \mathbf{x})$ via normalizing flows (Rezende and Mohamed, 2015). Following Manduchi et al. (2024) we use autoregressive normalizing flows, thereby minimizing the loss

$$\ell_{\text{NPE}}(\phi_s) = \frac{1}{N_s} \sum_{i=1}^{N_s} \log p_z \left(f_{\phi_s}(\theta_i; \mathbf{x}_i) \right) + \log \left| J_{f_{\phi_s}}(\theta_i; \mathbf{x}_i) \right| ,$$

where p_z denotes the isotropic Gaussian distribution and $J_{f_{\phi_s}}$ is the Jacobian's determinant of the function f_{ϕ_s}. Note that the function $f_{\phi_s}(\theta_i; \mathbf{x}_i)$ decomposes as $f_{\phi_s}(\theta_i; \mathbf{x}_i) = h_\tau(\theta_i; g_\eta(\mathbf{x}_i))$ where h_τ is invertible with respect to the first argument and parameterized via a three-step autoregressive normalizing flow, and g_η is a 1D-CNN encoder processing the APW segment into a 140-dimensional representation, used to condition the normalizing flow. The NPE model is trained for 50 epochs with a learning rate of 0.0003 and batch size 256, with early stopping on a validation set (20%). At inference time, we sample $K = 20$ values for the θ parameters from the normalizing flow

model for each of $M = 10$ APW generations obtained with the conditional VAE model (see also Section 2).

A.2.3. Baselines

In this section we describe implementation and training details for the baselines included in Section 3. We report three key baseline comparisons.

APW: We compare our approach with inference performed directly on ground truth VitalDB APWs, using the same NPE model trained on in-silico APWs that we use as part of our proposed approach. Note that in comparing with this baseline one should keep in mind that it requires invasive measurements. We train the model for 50 epochs with a learning rate of 0.0003 and early stopping on the validation set.

PPG Supervised: We also include as a baseline a supervised density estimator trained on a subset of VitalDB subjects. For consistency, we use the same architecture as the NPE model trained on in-silico APW that is part of our approach. We train the model on the labeled PPG data for 86 (two thirds) of the patients for training, and split the remaining ones equally between validation and set sets. To achieve this and prevent data leakage, we use subject-wise K-fold cross-validation, guaranteeing that every subject in the VitalDB dataset serves as part of a test set exactly once across the folds. The reported performance is then the aggregate over all these test sets, providing a fair comparison. We train the model for 50 epochs with a learning rate of 0.0003 and early stopping on the validation set.

PPG Windkessel: We train a density estimator to predict cardiovascular parameters on a dataset of in-silico PPGs (see Appendix A). As estimator we use the same NPE model architecture used for the other methods, and keep a 20% validation split for model selection. Performance is evaluated on the VitalDB dataset. Compared to the other models, where longer training did not result in an improvement, for this baseline we observed improved performance when training for a larger number of epochs so we report the results for the model trained for 250 epochs with a learning rate of 0.0003, with early stopping on the validation set.

A.3. Metrics and Evaluation

Following recent work (Manduchi et al., 2024) we focus our evaluation on assessing the ability of a model to predict temporal trends in cardiovascular parameters. To this end, we compare the predicted and ground truth values of a given cardiovascular biomarker for a given subject, over the time frame the subject is monitored. We compute the Sperman's correlation between the two time series. To mitigate short-term fluctuations and noise, both the predicted and ground truth time series are smoothed using an exponential moving average with a window size of 16 segments (i.e., 128 seconds), prior to computing correlation.

Appendix B. Additional qualitative and quantitative results

B.1. Generative modeling for PPG-to-APW mapping

Given the complexity of the cardiovascular system, the mapping between a PPG and the corresponding APW is complex and influenced by unobservable factors, introducing inherent ambiguities. Therefore, a single PPG segment may correspond to multiple physiologically plausible APWs. Consequently, training a deterministic model to map PPGs to APWs using a standard regression loss (e.g., Mean Squared Error) may be suboptimal, due to forcing a single averaged solution that fails to capture the variation in plausible generations. A generative model on the other hand may be a more sensible solution, with the possibility of capturing a distribution of plausible APWs through multiple generations. Therefore, in this work we employed a conditional VAE to model the conditional distribution of physiologically plausible APWs given an input PPG. To empirically validate this choice, in this section we compare the APW generations obtained with our approach, against the ones obtained with a neural network trained for direct APW reconstruction with an MSE loss.

We repeat the experiment described in Section 3, where produced APWs from VitalDB PPG inputs, are used for inference of cardiovascular parameters with our NPE model trained on in-silico APWs. Performance is assessed via the Spearman correlation between the ground truth and predicted values for HR, SV, CO and SVR. The results demonstrate a significant performance advantage for our generative model, confirming that a deterministic mapping lacks the expressive power to capture the physiological variations inherent in the PPG-to-APW relationship, thereby empirically validating our modeling choice.

Model	**HR**	**SV**	**CO**	**SVR**
Ours (1D-CNN)	0.94	0.31	0.40	0.24
Ours (ConditionalVAE)	0.96	0.49	0.53	0.29

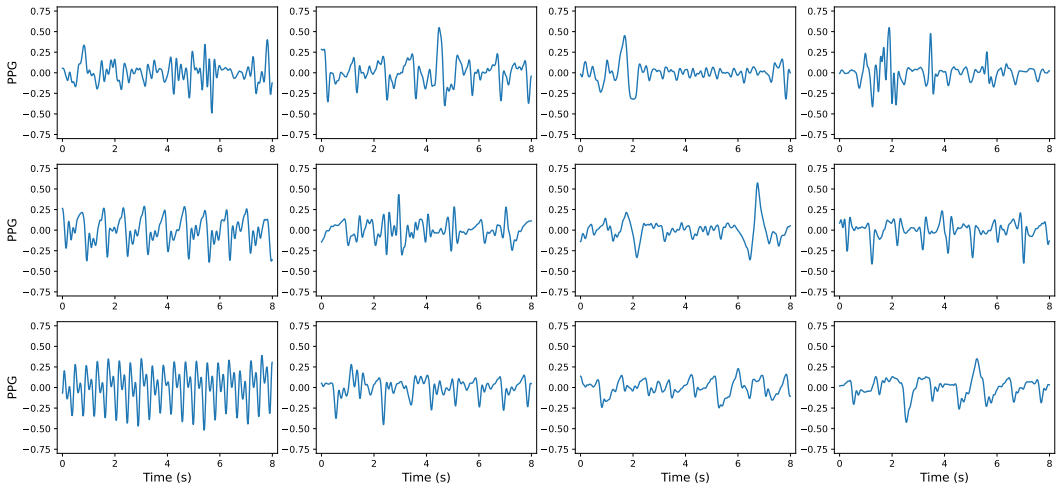
Table 1: Comparison between modeling the PPG-to-APW generation via a conditional VAE model vs a determinisitc 1-D CNN trained via MSE loss. Physiological relevance of the generated signals is assessed by evaluating the accuracy of cardiovascular parameter inference on the VitalDB dataset, using an NPE estimator trained in-silico. Average Spearman correlation between ground truth and predicted values per-subject for different cardiovascular parameters is reported.

Model	**HR**	**SV**	**CO**	**SVR**
APW	0.962 (0.002)	0.653 (0.006)	0.725 (0.006)	0.341 (0.008)
PPG Windkessel	0.804 (0.040)	0.348 (0.035)	0.438 (0.027)	0.289 (0.020)
PPG Supervised	0.967 (0.001)	0.190 (0.010)	0.333 (0.020)	0.246 (0.006)
Ours	0.966 (0.005)	0.489 (0.003)	0.531 (0.003)	0.288 (0.001)

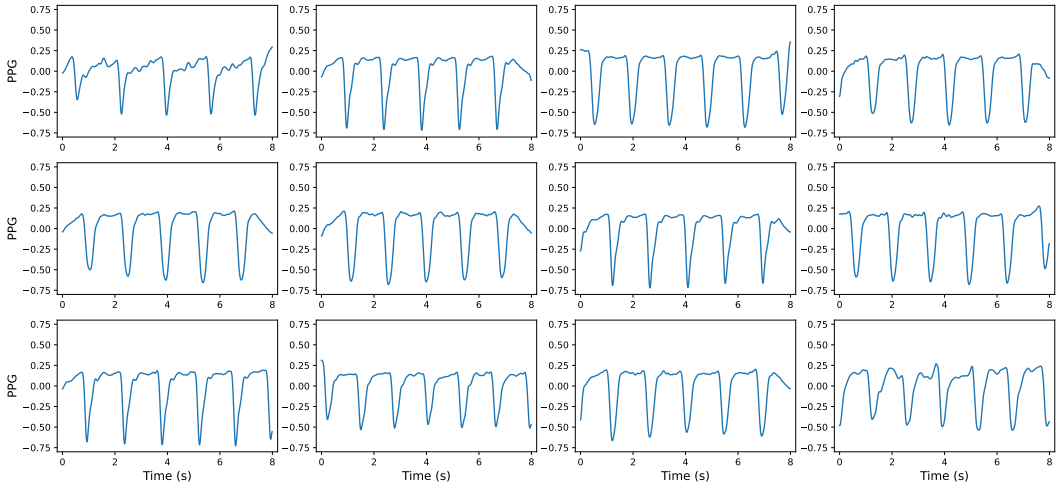
Table 2: Per-subject Spearman correlation between ground truth and predicted values for four cardiovascular biomarkers, obtained with the approaches compared in Figure 2. Results are averaged across three independent runs for each approach, and standard deviations are reported in parenthesis.

Model	**HR** (bpm)	**SV** (mL)	**SVR** ($dyn \cdot s/cm^5$)	**CO** (L/min)
APW	0.755	18.087	412.992	1.272
Ours	1.199	20.243	454.516	1.425
PPG Windkessel	2.951	20.960	557.478	1.552
PPG Supervised	0.725	18.789	328.499	1.329

Table 3: Average per-subject MAE results on the VitalDB dataset, for the models compared in Figure 2.



(a) High predicted uncertainty.



(b) Low predicted uncertainty.

Figure 4: Rejecting unusable samples via predicted uncertainty for HR inference. (a) Twelve VitalDB PPG segments with the highest predicted uncertainty. (b) Twelve VitalDB PPG segments with the least predicted uncertainty. All PPG signals underwent bandpass filtering.

B.2. Rejection of unusable PPG segments

In this section we demonstrate the utility of uncertainty modeling in our cardiovascular parameter inference, through the rejection of unusable PPG measurements. In detail, we use the uncertainty estimator associated with the cardiovascular parameter prediction (see Section 2). In Figure 4(a) we showcase segments with highest predicted heart rate (HR) uncertainty on the VitalDB dataset, demonstrating the ability to identify overly noisy input PPG samples for which inference is unreliable. Figure 4(b) complements these results, showing that low predicted uncertainty corresponds to clean PPG segments.

B.3. Absolute value prediction

Table 3 reports average per-subject Mean Absolute Error (MAE) for the compared models across the considered biomarkers. Our approach demonstrates superior performance to the PPG Windkessel model baseline, that also does not use labeled in-vivo data for training. However these results highlight that all approaches, including ours, currently exhibit signif-

icant limitations in absolute value prediction. This difficulty stems from the inherent dependence of complex hemodynamic parameters on unobservable, subject-specific physiological characteristics. Commercial devices often address this by integrating patient demographics (Hendy and Bubunek, 2016; Manecke, 2005) or requiring an initial gold-standard assessment (Hendy and Bubunek, 2016). In this work we focus on PPG-based cardiovascular assessment without resorting to in-vivo PPG signals with hemodynamic labels or patient-specific metadata for training or calibration. Consequently, within this specific setting, accurate absolute value prediction remains an elusive target. As discussed in section 3, personalization and calibration strategies represent promising directions to improve this performance aspect in future work.