Early Warning of In-Hospital Cardiac Arrest from Photoplethysmography Using Deep Residual Networks

Anonymous Author(s)

Affiliation Address email

Abstract

Early detection of in-hospital cardiac arrest remains a critical challenge for improving patient outcomes. We propose a deep learning framework that leverages continuous photoplethysmography (PPG) signals to predict cardiac arrest within a 24-hour window. We used the SCOPE dataset, a recently released collection of paired ECG and PPG waveforms from 4,517 ICU admissions across 3,785 patients at Seoul National University Hospital. We trained a residual 1D convolutional neural network on 5-minute PPG segments sampled at 125 Hz, and evaluated performance using patient-level stratified 5-fold cross-validation. The model achieved strong discrimination with both AUROC and AUPRC, demonstrating that PPG signals contain predictive signatures of impending deterioration. These findings highlight the feasibility of non-invasive waveform-based risk prediction, and position PPG monitoring as a promising biomarker in critical care.

1 Introduction

Cardiac arrest is among the most severe clinical emergencies, with survival rates below 25% despite advances in monitoring and resuscitation (1; 2; 3). Early identification of patients at risk is critical, as interventions before arrest significantly improve outcomes (4). Conventional early warning systems such as the Modified (MEWS) and the National Early Warning Score (NEWS) rely on manual and sparse vital signs and struggle to detect subtle changes that precede cardiac arrest (5). Although NEWS has stratified risk up to 24 hours prior to in-hospital cardiac arrest, its discriminative ability is limited (AUC of 0.58–0.64) (6). These limitations show the need for continuous, data-driven monitoring to detect physiologic deterioration earlier.

PPG, obtained through pulse oximetry, represents a promising signal for this task. PPG is non-invasive, low-cost, and already collected continuously in many hospital settings and wearable devices. Beyond heart rate and oxygen saturation, PPG captures beat-to-beat variability and morphological changes reflecting circulatory and autonomic status (7)(8). Prior studies have demonstrated associations between PPG features and cardiovascular instability (9), suggesting that PPG may contain early signs of clinical deterioration that are invisible to current early warning scores. Despite this, relatively little work has examined the use of continuous PPG time series for predicting in-hospital cardiac arrest.

Deep learning has transformed time-series analysis in healthcare, with CNNs and residual architectures showing strong performance on ECG, EEG, and other biosignals (10; 11; 12). Residual networks in particular allow efficient training of deep architectures while preserving temporal detail, making them well-suited for noisy physiological data. In this study, we investigate whether deep residual networks applied directly to raw PPG signals can provide early warning of in-hospital cardiac arrest.

36 2 Methods

We constructed a balanced experimental cohort by matching clinical outcomes with cached PPG recordings sampled at 125 Hz, filtering to patients with valid waveforms. To address class imbalance in the SCOPE dataset (71 cardiac arrest vs. >3,000 non-arrest), non-arrest cases were subsampled to achieve 1:1 and 1:2 ratios during training, while evaluation remained patient-level under the original distribution (Figure 1).

Continuous PPG recordings were segmented into five-minute windows, with or without 50% overlap. For cardiac arrest patients, only the last 24 h before the event were used; for non-arrest patients, only the first 24 h of admission. Segments were normalized (zero mean, unit variance; near-zero variance handled by mean subtraction) and missing values were replaced with zero. Each window received a binary label indicating whether it fell within the predictive horizon.

We implemented a 1D residual neural network in PyTorch with three residual blocks (16, 32, 64 channels) including convolution, batch normalization, and ReLU, using projection shortcuts when dimensions changed. The network ended with adaptive average pooling, a fully connected layer, and sigmoid activation to output the probability of cardiac arrest.

Models were trained for 10-20 epochs using Adam (lr = 1e-3), batch size 64, and binary cross-entropy loss. Evaluation followed patient-level five-fold stratified cross-validation, ensuring windows from the same patient remained in a single fold. Performance was assessed using AUROC and AUPRC on held-out patients.

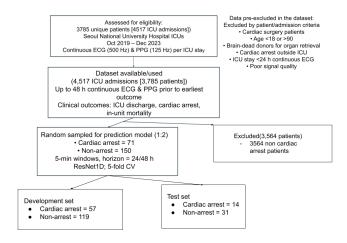


Figure 1: Flowchart of Study Cohort.

55 3 Results

Table 1: 24 Hr Horizon, 50% windows overlap, 1:1 dataset

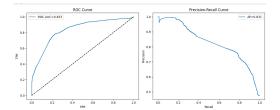
Stage	AUROC	AUPRC
Fold ₁	0.892	0.890
$Fold_2$	0.892	0.893
$Fold_3$	0,890	0.889
$Fold_4$	0,865	0.870
$Fold_5$	0.894	0.883
Mean	0.888 ± 0.011	0.885 ± 0.008

Fold-wise the best AUROC and AUPRC values are reported in Table 1. Table 1 shows the results to patient-level, five-fold stratified cross-validation with 24 hour horizon, 1:1 random sampled class

balance, and 50% overlap in windows. yields similar performance, suggesting robustness across different data splits.

Figure 2 presents the ROC and precision—recall curves for a 12-hour horizon, no overlap, and 1:2 cardiac arrest to non prediction. The model achieves strong discriminative performance (AUROC = 0.833) and maintains high precision (AUPRC = 0.831), despite the class imbalance. Figure 3 shows the corresponding curves for a 24-hour horizon, where ROC performance remains similar, but precision drops more notably at lower recall levels reflecting, as expected, increased uncertainty for longer-range predictions. Figure 4 displays the ROC curves stratified by fold for the 24-hour horizon, illustrating trends across cross-validation splits.

In additional (non-CV) sensitivity analyses, we varied the fixed train-test split (80/20 vs. 70/30), windowing strategy (300s/300s non-overlapping vs. 300s/150s overlapping), prediction horizon (24h vs. 48h), and class ratio (1:1 vs. 1:2, via down-sampling of non-arrest windows). These ablations yielded results that were qualitatively consistent with those from cross-validation. Among the variations tested, overlapping windows, a 1:1 class ratio, and a shorter prediction horizon (24h) led to modest improvements in both AUROC and AUPRC, indicating that these choices may enhance early warning performance.



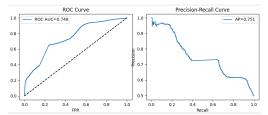


Figure 2: 12 Hr Horizon, No windows overlap, 1:2 dataset

Figure 3: 24 Hr Horizon, No windows overlap, 1:2 dataset.

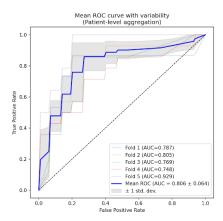


Figure 4: 24 Hr Horizon, 50% windows overlap, 1:2 dataset ROC curve full.

4 4 Discussion

Our findings demonstrate that deep residual networks trained on PPG segments can extract signals for early warning of in-hospital cardiac arrest. Importantly, evaluation was performed using patient-level cross-validation, which mitigates the optimistic bias that occurs when windows from the same patient appear in both training and validation. The stability of AUROC and AUPRC across folds suggests that the model is learning features that generalize beyond individual patients.

The dataset included only 74 cardiac arrest cases, yielding a more balanced cohort of approximately 300 samples after downsampling. While this approach reduced class imbalance, it may also have excluded patients and limited statistical power. We also constrained the prediction horizon

to 24 hours before arrest and limited control segments to the first 24 hours of admission. While this simplification aids model learning, it may not capture the full diversity of clinical trajectories, including gradual deterioration or recurrent instability.

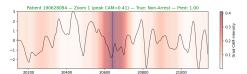
The current model also relies on a relatively shallow architecture and short fixed-length windows. More expressive architectures such as temporal transformers, diffusion models, or self-supervised pretraining approaches could capture longer-term dependencies and subtle temporal trends in the data. Finally, while AUROC and AUPRC provide useful measures of discrimination, clinical deployment would require additional evaluation of calibration, interpretability, and fairness across subgroups.

Future work should validate these methods in larger multi-institutional cohorts, explore multimodal integration with ECG or electronic health records, and investigate model interpretability techniques like saliency maps to identify physiologic correlates of risk. By grounding early warning in high-frequency physiological signals, this approach may complement existing early warning scores.

4.1 Grad-CAM

To better understand which waveform segments influenced the model's decisions, we applied Gradient-weighted Class Activation Mapping (Grad-CAM) to visualize temporal attribution across input PPG windows (15).

For the non-arrest case Figure 5, the model assigned moderate attention to regions with stable and smooth waveform morphology, suggesting that these segments contributed to the confident non-arrest prediction. In contrast, the arrest case Figure 6 shows distributed attention across the waveform, with higher attribution around irregular patterns, reflecting subtle physiologic instability captured in the PPG signal.



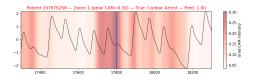


Figure 5: Non-arrest case Grad-CAM zoomed.

Figure 6: Arrest case Grad-CAM zoomed.

4.2 Related Works

Prior work on cardiac arrest prediction has explored a range of modalities and modeling strategies. Kataria et al. (13) developed PPG-GPT, a large pre-trained foundation model trained on extended waveform histories of up to 24 hours. Park et al. (14) incorporated multimodal data from the electronic medical record, including vital signs, laboratory results, and diagnostic codes, to achieve strong discriminative performance. Lee et al. (16) focused on heart rate variability features derived from ECG, emphasizing the predictive value of engineered physiological markers.

In contrast, our study uses a lightweight 1D ResNet trained directly on shorter, sliding-window PPG segments without pretraining, multimodal integration, or handcrafted features. We further extend prior work by systematically examining windowing strategies, prediction horizons, and class balancing methods, providing insights into how these design choices affect performance.

5 Conclusion

We presented a residual neural network for early warning of in-hospital cardiac arrest using only PPG time series. By segmenting continuous PPG recordings into fixed windows and restricting positive samples to the 24-hour pre-arrest horizon, we framed the task as a patient-level binary prediction problem. Training with five-fold stratified cross-validation demonstrated that the model could distinguish arrest from non-arrest cases using waveform-derived features alone. These results indicate that routinely collected PPG signals contain predictive information about clinical deterioration, and that residual networks can effectively extract these patterns. Future work will involve scaling to larger and more diverse datasets, incorporating additional modalities such as ECG or EHR-derived features, and extending to self-supervised representation learning for improved generalization.

5 References

- 126 [1] Abella, B.S. Not all cardiac arrests are the same. CMAJ. 2011;183(14):1572-1573. doi:10.1503/cmaj.110982
- [2] Andersen, L.W., Holmberg, M.J., Berg, K.M., Donnino, M.W. & Granfeldt, A. (2019) In-Hospital Cardiac Arrest: A Review. *JAMA – Journal of the American Medical Association* 321(12):1200-1210.
- [3] Ricceri, S., Salazar, J.W., Vu, A., Vittinghoff, E., Moffatt, E. & Tseng, Z.H. (2021) Factors
 Predisposing to Survival after Resuscitation for Sudden Cardiac Arrest. *Journal of the American College of Cardiology* 77(19):2353-2362.
- [4] Chugh, S.S. (2010) Early identification of risk factors for sudden cardiac death. *Nature Reviews Cardiology* **7**(6):318-326.
- [5] Mathukia, C., Fan, W.Q., Vadyak, K., Biege, C. & Krishnamurthy, M. (2015) Modified
 Early Warning System improves patient safety and clinical outcomes in an academic community hospital. *Journal of Community Hospital Internal Medicine Perspectives* 5(2):26716.
 doi:10.3402/jchimp.v5.26716.
- [6] Spångfors M, Molt M, Samuelson K. In-hospital cardiac arrest and preceding National Early
 Warning Score (NEWS): A retrospective case-control study. Clin Med (Lond). 2020;20(1):55-60.
 doi:10.7861/clinmed.2019-0137
- [7] Tamura, T. (2019) Current progress of photoplethysmography and SPO for monitoring and screening: a review. *Journal of Medical and Biological Engineering* **39**(6):845-851. doi:10.1007/s40846-019-00488-3.
- [8] Almarshad, M.A., Islam, M.S., Al-Ahmadi, S. & BaHammam, A.S. (2022) Diagnostic Features and Potential Applications of PPG Signal in Healthcare: A Systematic Review. *Healthcare* (*Basel*) **10**(3):547. doi:10.3390/healthcare10030547.
- [9] Karimpour, P., May, J.M. & Kyriacou, P.A. (2023) Photoplethysmography for the assessment of arterial stiffness. *Sensors (Basel)* **23**(24):9882. doi:10.3390/s23249882.
- [10] Petmezas, G., Stefanopoulos, L., Kilintzis, V., Tzavelis, A., Rogers, J.A., Katsaggelos, A.K.
 & Maglaveras, N. (2022) State-of-the-art deep learning methods on electrocardiogram data:
 systematic review. *JMIR Medical Informatics* 10(8):e38454. doi:10.2196/38454.
- 154 [11] Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T.H. & Faubert, J. (2019) Deep learning-based electroencephalography analysis: a systematic review. *Journal of Neural Engineering* **16**(5):051001. doi:10.1088/1741-2552/ab260c.
- 157 [12] Ganapathy, N., Swaminathan, R. & Deserno, T.M. (2018) Deep learning on 1-D biosignals: a taxonomy-based survey. *Yearbook of Medical Informatics* **27**(1):98. doi:10.1055/s-0038-159 1667083.
- [13] Kataria, S., Xiao, R., Ruchti, T., Clark, M.T., Lu, J., Lee, R.J., Grunwell, J. & Hu, X. (2025)
 Continuous cardiac arrest prediction in ICU using PPG foundation model. arXiv preprint arXiv:2502.08612.
- 163 [14] Park, H. & Park, C.S. (2025) A machine learning approach for predicting in-hospital car-164 diac arrest using single-day vital signs, laboratory test results, and International Classifi-165 cation of Disease-10 block for diagnosis. *Annals of Laboratory Medicine* **45**(2):209-217. 166 doi:10.3343/alm.2024.0315.
- 167 [15] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, "Grad-CAM:
 Visual Explanations from Deep Networks via Gradient-Based Localization," 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 618-626, doi: 10.1109/ICCV.2017.74.
- 171 [16] Lee, H., Yang, H.L., Ryu, H.G. et al. Real-time machine learning model to predict in-172 hospital cardiac arrest using heart rate variability in ICU. npj Digit. Med. 6, 215 (2023). 173 https://doi.org/10.1038/s41746-023-00960-2