

Differentially and Integrally Attentive Convolutional-Based Real-Time Photoplethysmographic Signal Quality Classification

Rafael Lima

Brazilian Development Bank

Italo Sandoval

`italo.sro@samsung.com`

Samsung R&D Institute Brazil

Arthur Valencio

Samsung R&D Institute Brazil

Maíssa Maniezzo

Samsung R&D Institute Brazil

Pedro Garcia

Universidade de Brasília

Article

Keywords:

Posted Date: September 18th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10514094/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: Competing interest reported. The authors are/were employed by Samsung Electronics, manufacturer of the devices being used for data acquisition.

Differentially and Integrally Attentive Convolutional-Based Real-Time Photoplethysmographic Signal Quality Classification

Rafael Lima¹, Italo Sandoval^{2,*}, Arthur Valencio², Maíssa Maniezzo², and Pedro Garcia Freitas³

¹National Bank for Economic and Social Development (BNDES), Rio de Janeiro/RJ, 20031-917, Brazil

²Samsung R&D Institute Brazil (SRBR), Campinas/SP, 13097-160, Brazil

³University of Brasília (UnB), Brasília/DF, 70910-900, Brazil

*italo.sro@samsung.com

ABSTRACT

Photoplethysmography (PPG) is a cost-effective optical technology that detects changes in blood volume within tissues, providing insights into the body's physiological dynamics over time. By analyzing PPG data as a time series, valuable information about cardiovascular health and other physiological parameters such as heart rate variability, peripheral oxygen saturation, and sleep stages can be estimated from commercial wearable devices. However, these indicators require reliable signals, which are often contaminated with movement noise, variable ambient lighting conditions, and other environmental effects. Hence, signal quality assessment is crucial in determining the trustworthiness of such data for wearable health applications. Modern approaches for PPG signal quality assessment are either based in the application of signal processing logical conditions or adoption of Convolutional Neural Networks. In this work, we propose the addition of differential and integral attention mechanisms composite with a two-stage rule and Convolutional-based procedure for the classification of segments. The approach is implemented in a way to balance storage size and classifier accuracy in the resulting models, increasing robustness across signals from different commercially available devices. Our proposed two-stage differential attention methods achieve accuracy values between 0.8952 and 0.9737, and F1-scores between 0.9183 and 0.9865 across four expert-annotated datasets.

Introduction

Photoplethysmography (PPG) is an essential sensing technique in wearable health monitoring systems, offering valuable insights into the cardiovascular system. It continuously provides physiological parameters that can be utilized to estimate health information such as heart rate, respiratory rate, and oxygen saturation, contributing to the comprehensive nature of these health-focused wearables. PPG signal plays a central role in these estimations due to its non-invasive nature and cost-effective implementation on edge devices¹. Its principle of operation is the emission of a light pulse onto the user's skin surface and the subsequent analysis of the intensity of light being reflected/transmitted, which varies proportionally to the blood volume flow in the biological tissue.

A major challenge impacting PPG's on-device and real-world performance is its susceptibility to noise, particularly motion artifacts², which can distort the signal's morphological properties and result in incorrect estimation of the derived physiological variables. Inaccurate assessments depending on these signals could potentially lead to severe consequences such as misdiagnoses, hence unreliable performance would be inadequate to real-world applications. Thus, Signal Quality Assessment (SQA) techniques are employed to identify the presence of noisy signal sections, marking these as unsuitable to critical clinical applications. The implementation of this techniques is usually through the employment of a Signal Quality Classifier (SQC) module, which may adopt rule-based approaches based on signal processing conditions or machine-learning models such as neural networks to differentiate high-quality signal segments from low-quality ones.

Significant progress has been made in recent years on the development of SQC modules³. For example, Elgendi⁴ proposed a technique employing indices to assess quality based on signal processing features and data statistical distributions. Selvaraj⁵ presented a statistical method involving the calculation of kurtosis and Shannon Entropy to identify motion artifacts and noise contained in PPG data. Li and Clifford⁶ introduced an alternate statistical approach, leveraging dynamic time warping to stretch each heartbeat for alignment with a dynamic template, with some of the derived features being associated with signal quality. Regarding frequency domain, Sivanjaneyulu et al.⁷ proposes a classification method making use of Deep Fourier Magnitude Spectrum. In the classification phase of the approach described by Li and Clifford⁶, a multi-layer perceptron type of neural

network is utilized to comprehend the correlations among parameters in the context of both high- and low-quality pulses.

Sun et al.⁸ introduced an approach that exploits the morphological features of the signal for evaluating its quality. Li et al.⁹ identified four waveform characteristics to evaluate signal quality through the application of a decision tree. Similarly, Sukor et al.¹⁰ utilized a basic decision-tree classifier to determine, with pulse-by-pulse precision, whether a particular pulse is suitable for use or not. Naeini et al.¹¹ was the first to adopt Convolution Neural Networks (CNN) architectures to categorize the signal into ‘reliable’ or ‘unreliable’ categories. Recently, Freitas et al.^{12,13} introduced a SQA technique that transforms PPG signals into two-dimensional representations and subsequently employs a vision transformer to evaluate their quality. Lucafo et al.¹⁴ proposed an ensemble of local magnitude comparison-based feature descriptors for subsequently feeding into a linear classifier.

Additionally, the work by Silva et al.¹⁵ explored the usage of attention-based approaches of Bahdanau et al.¹⁶, Luong et al.¹⁷, and Vaswani et al.¹⁸ in conjunction with a Convolutional-based strategy to find an efficient Signal Quality Classifier (SQC) under strict trade-offs of memory, classification performance, and power consumption, similar to those observed in commercial wearables.

In the present work, we enhance the performance and efficiency of such CNN-based approach to perform PPG SQA by jointly leveraging differential and integral attention layers¹⁹ prior to the final dense layer. This way, we develop an efficient SQC capable of being deployed in real-time Health Monitoring Applications (HMA) of resource-constrained embedded devices. The new machine learning architecture provided by such layers remove the common mode attention noise, optimizing the performance of subsequent classifier mechanisms.

Methods

Overview

The objective of this work is to develop and compare models that perform binary quality classification of real-time unidimensional PPG signal segments obtained from wearable devices. Our aim is to enhance the capabilities of a compact CNN using a differential attention¹⁹ mechanism for quality classification of PPG sensor data, a functionality essential for wearable applications such as heart rate and sleep assessment.

CNNs were originally proposed for computer vision tasks, automatically extracting local and global features from images or frames for classification²⁰. However, they have expanded to other domains, including time series classification, such as biomedical ones²¹. Meanwhile, attention mechanisms originated in natural language processing to overcome the issue of vanishing gradients in long-term dependency learning of sequential models¹⁶, such as Recurrent Neural Networks (RNNs). They have achieved state-of-the-art performance across various domains, including Large Language Models (LLMs)²² and vision-related tasks^{12,13}.

We utilize Neural Architecture Search (NAS)²³ techniques and attention mechanisms to identify neural network solutions that are lightweight yet able to deliver competitive performance within strict memory and computational constraints, such as the architecture shown in Figure 1. Our baseline CNN, detailed in the Convolutional Backbone module of Figure 1, was discovered through a NAS process analogous to the one described by Lima et al.²⁴, but specifically tailored for PPG signal quality assessment rather than systolic peak classification as in the original work. The details of the processing steps, the definitions of the Differential Attentive and Differential-Integral Attentive (DINT Attentive) Layers, and the experimental setup enabling the generation of the datasets adopted for the comparative analysis, are discussed in the following subsections.

Preprocessing

Regarding the preprocessing used in this work, we start from the discrete unidimensional PPG vector

$$Y = \{y_k\}_0^{N_{samples}-1} \in \mathbb{R}^{N_{samples}}, \quad (1)$$

collected with a fixed sampling frequency $f_s=25\text{Hz}$.

This PPG vector is subject to a windowing operation, resulting in segments of 9 seconds with 3 seconds overlap on each edge. For each segment, it is applied an order-2 zero-phase Butterworth bandpass filter with cutoff frequencies of 0.8 and 4.5Hz. The lower cutoff frequency aims to remove the baseline wandering while the higher cutoff frequency removes environmental noise.

The PPG segments are, then, further converted to a dataset $\mathcal{Y} \in \mathbb{R}^{N \times L}$ composed of N non-overlapping segments of 3-seconds with signal length of L samples, in the form

$$\mathcal{Y} = \{Y_k\}_{k=0}^{N-1}, \text{ with } Y_k = \{y_{N \cdot k + j}\}_{j=0}^{L-1}, \quad (2)$$

with $L = 3 \cdot f_s = 75$ samples, and $N = \lfloor \frac{N_{samples}}{L} \rfloor$, for $N_{samples}$ the total number of samples.

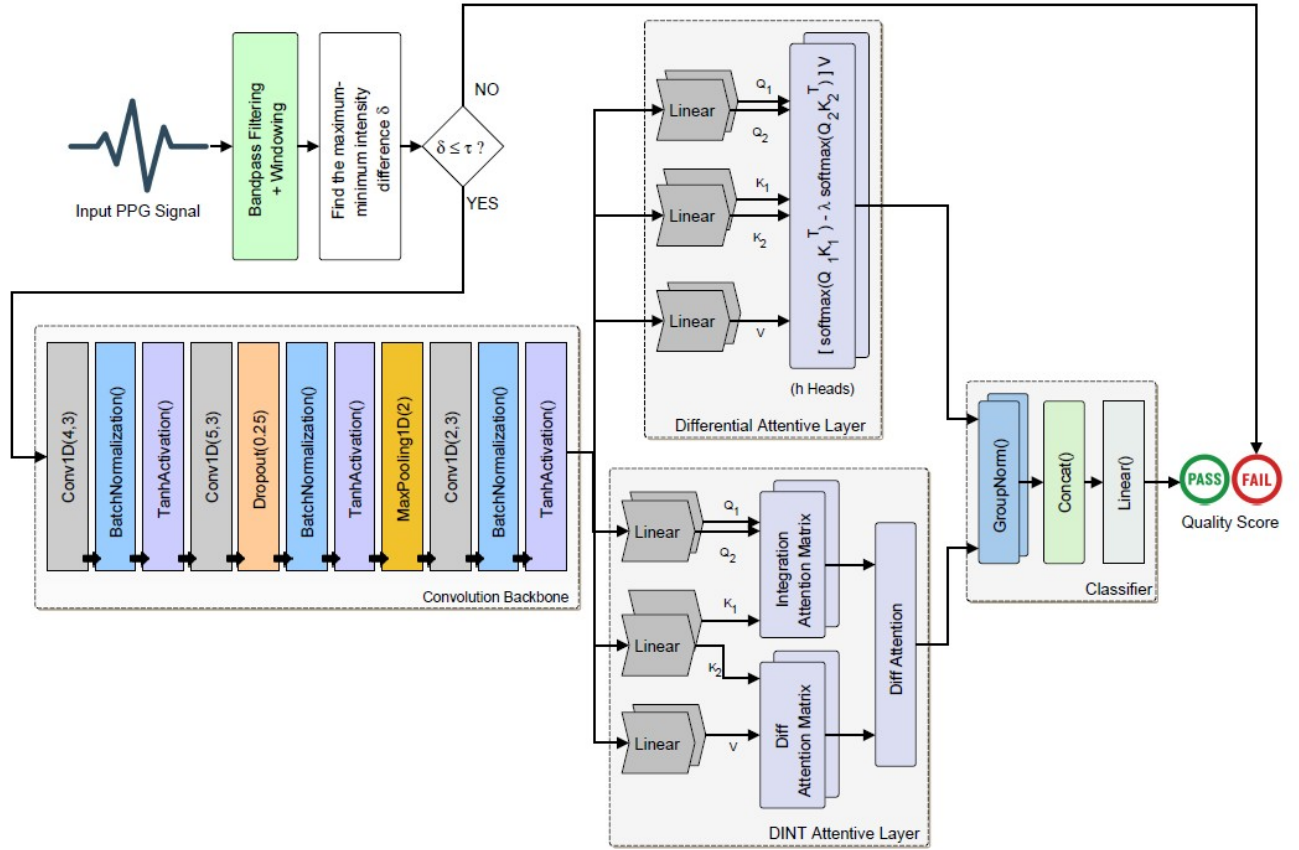


Figure 1. Overview of the proposed PPG classification method

A binary quality label needs to be associated with each PPG segment, with 0 meaning low-quality (unreliable) PPG data and 1 meaning high-quality (reliable) PPG data. To construct this, a cardiologist provides a sample-by-sample ground truth of zeros-and-ones labels for the entire dataset. The label of a 3-seconds (= 75 samples) segment is, then, computed by evaluating the prevalence of samples labelled as 1. If it is equal or higher than 50% of the segment samples, then the segment is considered as having a label 1. Otherwise, the segment label is assigned to 0.

Two-Stage Signal Classification

A Two-Stage Signal Quality Classification (SQC) pipeline is here proposed, inspired by the work of Lucafo et al.²¹. It has a hierarchical structure composed of an initial rule-based decision stage with the purpose of discarding PPG windows that are most likely to be of low quality based on the empirical observation, from training data, of the signal amplitude ranges generally considered invalid due to high motion or other external factors corrupting the PPG signal. For this, the amplitude difference δ is compared to a fixed threshold τ , i.e., if $\delta < \tau$, the solution proceeds for a second stage of classification that investigates the morphology of the signal to characterize its signal quality. Such second stage of the SQC consists in classifying the PPG windows by a machine-learning solution combining CNNs, Attention mechanisms and a Linear Regressor Classifier, as previously described in Fig. 1.

In this work, we investigate the impact of this two-stage classification on three state-of-the-art attention-based PPG classification methods: Additive Attention Layer (AAL), Dot Product Attention Layer (DPAL) and Non-Scaled Dot Product Attention Layer (NSDPAL), described by Silva et al.¹⁵, besides the composite classifiers Differential Attention and Differential-Integral Attention Layers, described in the next subsections.

Differential Attention

The Differential (DIFF) attention mechanism¹⁸ maps query, key, and value vectors to the final outputs, by using query and key vectors to compute intermediate attention scores, which are then used to compute a final weighted sum of value vectors. The proposed design adopts paired softmax functions for noise canceling in the intermediate attention scores. The final TanhActivation layer of the Convolutional Backbone produces vectors $X \in \mathbb{R}^{N \times d_{model}}$, with d_{model} being the hidden dimension

of the model, which are considered inputs to the differential attention mechanism. Such inputs are initially projected to corresponding query vectors $Q_1, Q_2 \in \mathbb{R}^{N \times d}$, key vectors $K_1, K_2 \in \mathbb{R}^{N \times d}$, and value vectors $V \in \mathbb{R}^{N \times 2d}$. The dimension vector d is given by $d = d_{model}/2h$, with h the number of attention heads, here set to $h = 1$. Afterwards, the differential attention operator $\text{DiffAttn}(\cdot)$ calculates the respective outputs from:

$$[Q_1; Q_2] = XW^Q, [K_1; K_2] = XW^K, V = XW^V \quad (3)$$

$$\text{DiffAttn}(X) = \left(\sigma \left(\frac{Q_1 K_1^T}{\sqrt{d}} \right) - \lambda \sigma \left(\frac{Q_2 K_2^T}{\sqrt{d}} \right) \right) V, \quad (4)$$

where $\sigma(\cdot)$ is the softmax function, defined by

$$\sigma(z_i) = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}, \text{ for } i = 1, 2, \dots, K \quad (5)$$

and $W^Q, W^K, W^V \in \mathbb{R}^{d_{model} \times 2d}$ are trainable parameters, and λ is a learnable scalar value. With the purpose of synchronizing the learning dynamics, the scalar λ is reparameterized as:

$$\lambda = e^{(\lambda_{q1} \cdot \lambda_{k1})} - e^{(\lambda_{q2} \cdot \lambda_{k2})} + \lambda_{init} \quad (6)$$

where $\{\lambda_{q1}, \lambda_{k1}, \lambda_{q2}, \lambda_{k2}\} \subset \mathbb{R}^d$ are learnable vectors, and $\lambda_{init} \in [0, 1]$ is a constant used for the initialization of λ . The best value of λ_{init} was empirically searched along the range of $[0, 1]$, reporting each tested value in the metric tables. The experimental results consider the variation of performance along different values of initialization λ_{init} .

Thus, the Differential attention mechanism computes the difference between two softmax attention functions, as a way of eliminating attention noise. The proposed idea can be shown to be analogous to that of differential amplifiers²⁵, more commonly studied in the electrical engineering domain, in which the difference between two signals being taken as output, as a way to remove the common-mode noise of the input.

Differential Integral Attention

Differential Integral (DINT) attention²⁶ is an extension of the differential (DIFF) attention which introduces an additional integral mechanism, that enhances the model's capability of extracting globally relevant information while retaining numerical stability through the normalization of rows in the corresponding final attention matrix. The attention matrix A_1 of the signal is calculated using Q_1 and K_1 :

$$A_1 = \sigma \left(\frac{Q_1 K_1^T}{\sqrt{d}} \right) \quad (7)$$

The integral mechanism computes the global attention importance scores through averaging the signal attention weights across each column:

$$G = \frac{1}{N} \sum_{m=1}^N A_1[m, n], \quad (8)$$

with $G \in \mathbb{R}^{1 \times N}$ then being expanded for matching the dimensions of the differential mechanism:

$$G_{expanded} = \text{repeat}(G, N) \quad (9)$$

with $G_{expanded} \in \mathbb{R}^{N \times N}$ being constructed by repeating G along N rows. Thus, the full Differential-Integral attention mechanism, represented by the DINT operator, calculates the output from the expression:

$$\text{DINT}(X) = (A_{diff} + \gamma G_{expanded}) V, \quad (10)$$

with γ being a scalar, A_{diff} being the differential attention mechanism, and $G_{expanded}$ being the expanded matrix of global importance scores. Usually, λ and γ are set to equal values for ensuring that the resulting attention matrix A_{final} has rows

summing up to 1. This normalization of rows assures numerical stability and consistency along the model, what increases data stability across layers. This unified configuration follows the method of parameterization used in the original DIFF Transformer, as a means of further increasing stability during training.

In the present work, we propose enhancing one-stage and two-stage CNN-based PPG classification²¹ by adopting both Differential and Differential-Integral attention mechanisms replacing the GlobalAveragePooling1D layer commonly employed between the Convolutional Backbone and a Classifier step. In the results section, we refer to the two composite classifiers as Differential Dot-Product Attention Layer (DFPAL) and Differential-Integral Dot-Product Attention Layer (DINTAL), respectively.

Experimental Setup

To assess the performance of the proposed architectures, we compare the identified neural network classifiers against the state-of-the-art SQC method using the same evaluation process on four distinct datasets:

- GW5: Comprising 149 sessions lasting approximately 35 minutes each, with 119 utilized for training and 30 for testing. PPG samples were obtained with Samsung Galaxy Watch 5, from different subjects at rest while sitting in a comfortable chair in a quiet room.
- GW6: Consisting of 94 sessions lasting around 35 minutes each, with 75 dedicated to training and 19 assigned to testing. PPG readings were collected utilizing a Samsung Galaxy Watch 6, from different subjects at rest while sitting in a comfortable chair in a quiet room.
- GW7: Including 37 sessions lasting approximately 35 minutes each, with 30 allocated for training and 7 designated for testing. PPG readings were acquired using a Samsung Galaxy Watch 7, from different subjects at rest while sitting in a comfortable chair in a quiet room.
- RING: Comprised of 73 sessions lasting around 35 minutes each, with 59 reserved for training and 14 set aside for testing. PPG readings were obtained using a Samsung Galaxy Ring, from different subjects at rest while sitting in a comfortable chair in a quiet room.

For each Galaxy Watch subject, the PPG sensor was carefully positioned on the wrist over the radial artery to optimize signal acquisition. For Galaxy Ring subjects, the device was inserted in the finger which provided the best adherence to the skin. The quality of the signals in the datasets were manually labeled by a cardiologist expert, taking into account the waveform characteristics and the dataset’s purpose of measuring Interbeat Intervals (IBIs) between PPG systolic peaks.

The PPG signals were partitioned into non-overlapping 3-second windows, each containing 75 samples. To prepare for the learning phase, each segment received a ‘1’ or ‘0’ quality label depending on whether the proportion of high-quality samples exceeded a given threshold: segments with more than or equal to 50% high-quality signal were classified as ‘reliable’ (‘1’), while others were deemed ‘unreliable’ (‘0’).

For the experiments, the data was split into training and test sets. Validation subsets were randomly selected from the training set. During training, we utilized data and corresponding labels from 64% of the subjects in each assessment group. For validation, we chose data from 16% of the subjects in each group. The remaining 20% of subjects in each group were reserved for testing the proposed methods and the entire pipeline. The signal segments used in the experiments were organized according to the subject they belonged to. This ensured that there was no overlap between training and testing sets, preventing potential training biases from influencing testing results. For our proposed method, we concurrently performed an ablation of the λ_{init} parameters across the range of values in {0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.99}.

For implementing the DL models, we used Keras 2.9.0 and Python 3.8.17. For the remaining classifiers, we used scikit-learn 1.3.0. All models underwent 5 trials of 100 training epochs each, with the resulting test set performance metrics being averaged across all trials. To gauge the performance of both the proposed and state-of-the-art methods, we compared the predicted quality indices with the pre-labeled indices provided in the benchmark dataset using Accuracy, Precision, Recall, F1-Score, Area Under the receiver operating characteristic Curve (AUC), Proportion of the signal classified as high-quality (Coverage), Matthews Correlation Coefficient (MCC), and Cohen’s Kappa metrics²⁷. With the exception of Coverage, all metrics are better when approaching to 1. Coverage is an auxiliary metric that supports informing whether the model is too strict (almost all real-world signal considered low-quality, resulting in coverage close to zero), or too permissive (all real-world signal considered acceptable, resulting in coverage close to 1).

Results

We have implemented 10 models adopting additive attention (AAL), dot-product attention (DPAL), non-scaled dot-product attention (NSDPAL), differential dot-product attention (DFPAL) and differential-integral dot-product attention (DINTAL)

mechanisms in Attentive-CNN or Two-Stage architectures. Then, their performance was compared against 6 benchmarks based on CNNs²¹ (Lucafo1 and Lucafo 2), descriptor-based features²⁸ (CASLBP SGD, CASLTP SGD and LBP SGD) and sets of signal processing rules²⁹ (Hao & Bo), considering the cardiologist label as the ground-truth.

The average performance metrics for our proposed method with DFPAL, as well as baseline out-of-the-box classifiers and state-of-the-art attention-based PPG classification methods, are shown in Table 1. It can be observed that architectures adopting DINTAL and DFPAL mechanisms generally outperform other methods in terms of accuracy, F1-score, AUC, MCC and Cohen's Kappa. In particular, the DINTAL mechanism resulted in the best performing accuracy on three datasets (GW5, GW7 and Ring), while the DFPAL has the best accuracy only on GW6 dataset. Regarding F1-scores, DINTAL is also the best performing in the same three datasets, while AAL is the best alternative for GW6. With both metrics, two-stage approaches show the best performance, with a notable exception for the GW5 dataset, in which the one-stage approach (without an initial rule-based selection by the signal amplitude) is preferred.

Regarding model size, Table 2 shows that the required increase in the model total size of the proposed composite DFPAL and DINTAL methods is not proportionally substantial when compared to the attention-based state-of-the-art comparison methods. The increase in model size is significant when compared with lower-performing methods, such as the models LBP, CASLBP and CASLTP, based on feature descriptors (unidimensional versions of local binary and local ternary patterns, techniques imported from image processing). When memory and computational burden restrictions are too strict, with a available memory cap smaller than 200KB, such memory performance trade-offs may have to be taken into consideration. If the available memory exceeds 451KB, the focus can be strictly on performance.

Analysing the impact of the free model initialization parameter λ_{init} on the test set, Figures 2, 3, 4, and 5 show how the performance metrics vary with the adopted value, for each dataset. Considering how variation reflect in the y-axis scales, the values of accuracy, precision, recall and F1-score are generally stable, while AUC, MCC and Cohen's Kappa may change substantially. Two-Stage architectures adopting DFPAL and DINTAL mechanisms are considerably more sensitive to λ_{init} selection than Attention-CNN architectures. Hence, fine-tuning step can be critical for the desirable high performance in such cases.

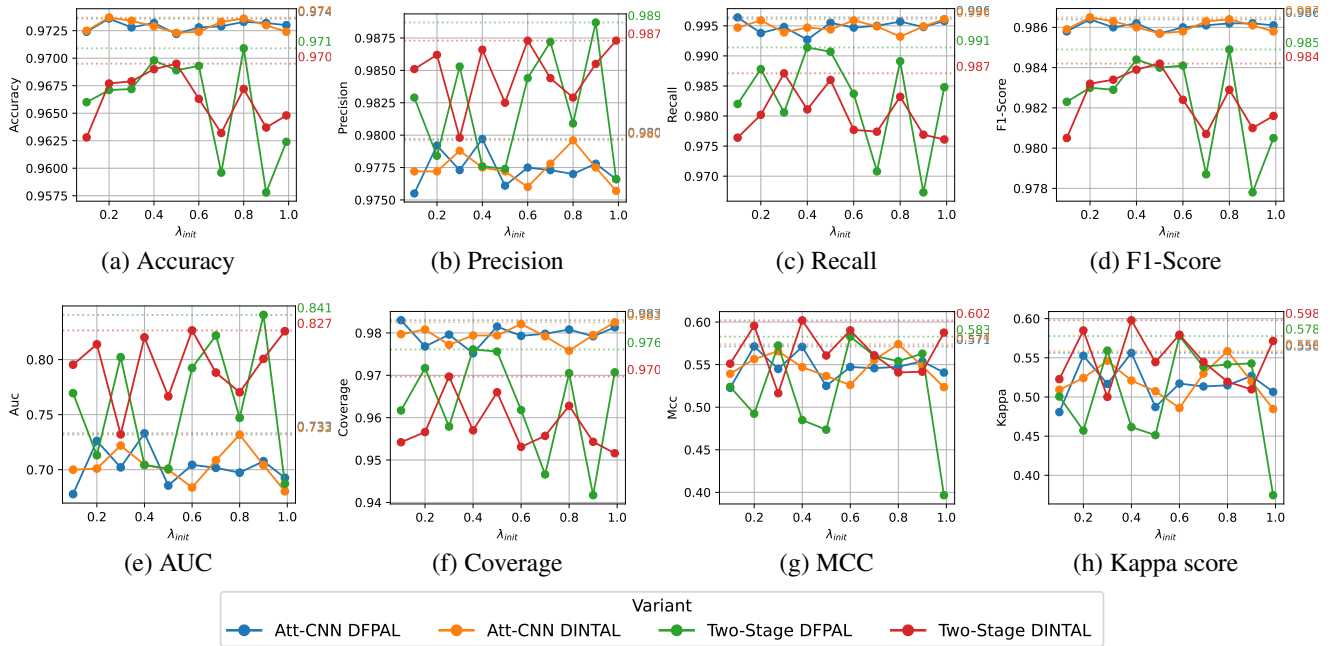


Figure 2. Performance comparison of DFPAL and DINTAL, with and without Two-Stage Classification pipeline, across several values of λ_{init} on GW5 PPG dataset. All the methods were trained for 5 trials and 1000 epochs. The test set performance metrics were then averaged across all trials.

Discussion

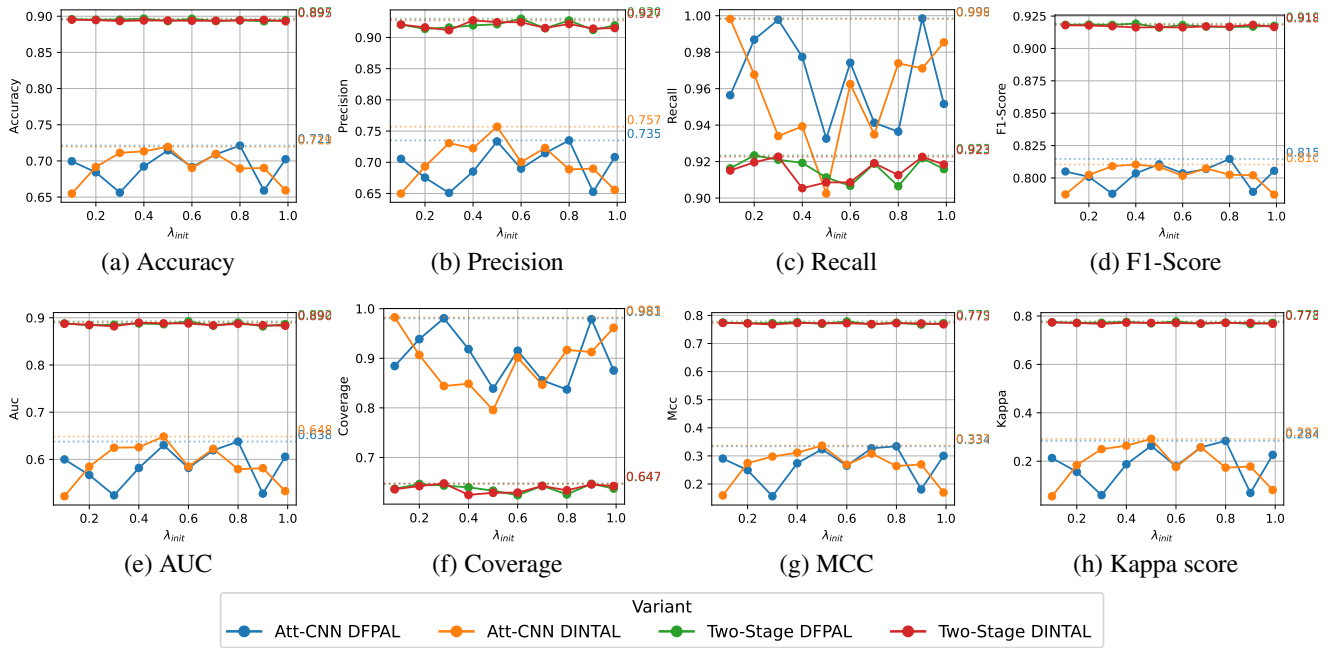
Wearable health monitoring apps benefit from PPG technology due to its simple configuration, convenience of use, low-complexity, and cost-effectiveness. Despite these advantages, PPG signals are vulnerable to significant degradation due to

Table 1. Performance comparison of the architectures after training for five trials (initial states) over 1000 epochs. Test metrics for each architecture are the average across the five outcomes from each trial.

Dataset	Model	Accuracy	Precision	Recall	F1-Score	AUC	Coverage	MCC	Kappa
GW5	Att-CNN+DFPAL	0.9736	0.9797	0.9964	0.9864	0.7330	0.9830	0.5714	0.5562
	Att-CNN+DINTAL	0.9737	0.9796	0.9961	0.9865	0.7318	0.9825	0.5740	0.5584
	Two-Stage+DFPAL	0.9709	0.9887	0.9914	0.9849	0.8406	0.9761	0.5829	0.5778
	Two-Stage+DINTAL	0.9695	0.9873	0.9871	0.9842	0.8265	0.9697	0.6020	0.5979
	Att-CNN+AAL	0.9722	0.9765	0.9951	0.9857	0.6904	0.9808	0.5285	0.4937
	Att-CNN+DPAL	0.9732	0.9783	0.9942	0.9862	0.7148	0.9781	0.5582	0.5352
	Att-CNN+NSDPAL	0.9734	0.9791	0.9936	0.9863	0.7249	0.9767	0.5676	0.5500
	Two-Stage+AAL	0.9707	0.9803	0.9895	0.9848	0.7396	0.9715	0.5462	0.5325
	Two-Stage+DPAL	0.9632	0.9816	0.9803	0.9808	0.7523	0.9614	0.5124	0.4911
	Two-Stage+NSDPAL	0.9708	0.9801	0.9897	0.9849	0.7376	0.9719	0.5466	0.5333
	Lucafo1	0.9720	0.9794	0.9918	0.9855	0.7284	0.9746	0.5526	0.5398
	Lucafo2	0.9677	0.9778	0.9890	0.9833	0.7056	0.9736	0.4866	0.4622
	CASLBP SGD	0.9647	0.9710	0.9930	0.9819	0.6174	0.9841	0.3584	0.3238
	CASLTP SGD	0.9631	0.9643	0.9986	0.9812	0.5260	0.9967	0.0774	0.0711
	LBP SGD	0.9638	0.9700	0.9932	0.9814	0.6033	0.9854	0.2944	0.2740
	Hao & Bo	0.5808	0.9702	0.5823	0.7278	0.5619	0.5776	0.0477	0.0209
GW6	Att-CNN+DFPAL	0.7212	0.7351	0.9985	0.8147	0.6378	0.9806	0.3342	0.2841
	Att-CNN+DINTAL	0.7194	0.7571	0.9982	0.8104	0.6484	0.9825	0.3367	0.2922
	Two-Stage+DFPAL	0.8967	0.9301	0.9234	0.9192	0.8925	0.6468	0.7786	0.7776
	Two-Stage+DINTAL	0.8952	0.9274	0.9227	0.9183	0.8897	0.6475	0.7733	0.7733
	Att-CNN+AAL	0.6915	0.7033	0.9511	0.7996	0.5909	0.8857	0.2703	0.1910
	Att-CNN+DPAL	0.6598	0.6563	0.9832	0.7871	0.5345	0.9583	0.1713	0.0853
	Att-CNN+NSDPAL	0.6891	0.6911	0.9691	0.8017	0.5806	0.9110	0.2528	0.1737
	Two-Stage+AAL	0.8967	0.9183	0.9210	0.9194	0.8872	0.6419	0.7765	0.7755
	Two-Stage+DPAL	0.8931	0.9158	0.9174	0.9165	0.8837	0.6409	0.7684	0.7679
	Two-Stage+NSDPAL	0.8962	0.9172	0.9210	0.9190	0.8866	0.6423	0.7748	0.7745
	Lucafo1	0.7006	0.7069	0.9427	0.8035	0.6068	0.8657	0.2847	0.2318
	Lucafo2	0.8960	0.9266	0.9099	0.9179	0.8906	0.6284	0.7767	0.7758
	CASLBP SGD	0.8487	0.8709	0.8965	0.8835	0.8302	0.6586	0.6688	0.6681
	CASLTP SGD	0.6396	0.6396	1.0000	0.7802	0.5000	1.0000	0.0000	0.0000
	LBP SGD	0.8545	0.8794	0.8954	0.8873	0.8386	0.6514	0.6824	0.6821
	Hao & Bo	0.4109	0.9395	0.0844	0.1549	0.5374	0.0575	0.1542	0.0552
GW7	Att-CNN+DFPAL	0.9555	0.9551	1.0000	0.9762	0.7473	0.9942	0.6738	0.6325
	Att-CNN+DINTAL	0.9533	0.9547	0.9996	0.9750	0.7444	0.9709	0.6569	0.6202
	Two-Stage+DFPAL	0.9623	0.9765	0.9895	0.9796	0.8625	0.9333	0.7405	0.7356
	Two-Stage+DINTAL	0.9626	0.9752	0.9887	0.9797	0.8551	0.9327	0.7465	0.7416
	Att-CNN+AAL	0.9532	0.9526	0.9985	0.9750	0.7325	0.9589	0.6527	0.6044
	Att-CNN+DPAL	0.9521	0.9520	0.9980	0.9745	0.7290	0.9590	0.6444	0.5982
	Att-CNN+NSDPAL	0.9503	0.9491	0.9993	0.9735	0.7117	0.9633	0.6276	0.5694
	Two-Stage+AAL	0.9599	0.9759	0.9805	0.9782	0.8600	0.9191	0.7380	0.7367
	Two-Stage+DPAL	0.9608	0.9677	0.9902	0.9788	0.8176	0.9361	0.7255	0.7153
	Two-Stage+NSDPAL	0.9607	0.9754	0.9819	0.9786	0.8580	0.9209	0.7416	0.7393
	Lucafo1	0.9536	0.9616	0.9888	0.9750	0.7822	0.9407	0.6750	0.6574
	Lucafo2	0.9607	0.9698	0.9879	0.9788	0.8284	0.9320	0.7290	0.7201
	CASLBP SGD	0.9295	0.9297	0.9984	0.9628	0.5938	0.9824	0.3971	0.2901
	CASLTP SGD	0.9231	0.9231	0.9993	0.9596	0.5517	0.9905	0.1896	0.1535
	LBP SGD	0.9296	0.9292	0.9992	0.9629	0.5904	0.9838	0.3580	0.2760
	Hao & Bo	0.6383	0.9302	0.6536	0.7678	0.5634	0.6428	0.0739	0.0518
RING	Att-CNN+DFPAL	0.8794	0.8761	0.9940	0.9279	0.7381	0.9017	0.6105	0.5707
	Att-CNN+DINTAL	0.8784	0.8760	0.9934	0.9275	0.7372	0.9013	0.6082	0.5666
	Two-Stage+DFPAL	0.9030	0.9479	0.9349	0.9380	0.8663	0.7803	0.7154	0.7152
	Two-Stage+DINTAL	0.9045	0.9494	0.9374	0.9390	0.8679	0.7824	0.7183	0.7181
	Att-CNN+AAL	0.8776	0.8706	0.9918	0.9272	0.7260	0.8947	0.6043	0.5526
	Att-CNN+DPAL	0.8764	0.8696	0.9912	0.9264	0.7240	0.8950	0.6004	0.5500
	Att-CNN+NSDPAL	0.8794	0.8723	0.9915	0.9281	0.7306	0.8924	0.6115	0.5632
	Two-Stage+AAL	0.8936	0.9492	0.9134	0.9308	0.8673	0.7555	0.7033	0.7003
	Two-Stage+DPAL	0.8958	0.9412	0.9252	0.9331	0.8568	0.7718	0.6990	0.6980
	Two-Stage+NSDPAL	0.8917	0.9471	0.9133	0.9297	0.8631	0.7572	0.6973	0.6941
	Lucafo1	0.8836	0.8787	0.9883	0.9302	0.7446	0.8831	0.6262	0.5860
	Lucafo2	0.8994	0.9417	0.9297	0.9355	0.8592	0.7753	0.7081	0.7066
	CASLBP SGD	0.8424	0.8602	0.9545	0.9048	0.6936	0.8713	0.4753	0.4528
	CASLTP SGD	0.8148	0.8218	0.9821	0.8935	0.5930	0.9421	0.2171	0.2110
	LBP SGD	0.8492	0.8704	0.9493	0.9081	0.7164	0.8563	0.5068	0.4919
	Hao & Bo	0.6180	0.8368	0.6379	0.7239	0.5917	0.5985	0.1536	0.1394

Table 2. Model Size Comparison across proposed attention-based solutions and state-of-the-art comparison methods.

Model Architecture	Model Size (Bytes)
Att-CNN+DFPAL	430382
Att-CNN+DINTAL	450862
Two-Stage+DFPAL	430512
Two-Stage+DINTAL	450992
Att-CNN+AAL	379182
Att-CNN+DPAL	368942
Att-CNN+NSDPAL	368942
Two-Stage+AAL	379312
Two-Stage+DPAL	369072
Two-Stage+NSDPAL	369072
Lucafo1	348462
Lucafo2	348544
CASLBP SGD	82222
CASLTP SGD	92462
LBP SGD	143662

**Figure 3.** Performance comparison of DFPAL and DINTAL, with and without Two-Stage Classification pipeline, across several values of λ_{init} on GW6 PPG dataset. All the methods were trained for 5 trials and 1000 epochs. The test set performance metrics were then averaged across all trials.

various factors, primarily motion artifacts, which can distort the waveform morphology, impacting subsequent signal analysis. Errors resulting from unreliable signals are unacceptable for HMA, as these can impact diagnoses and preventive medical care. Thus, assessing PPG signal quality is essential to avoid misinterpretation, reduce false positive/negative rates in downstream solutions, and differentiate between reliable and noisy signals. To address this, in this study we leveraged ML models to develop robust and high-performing models for classifying PPG segments with the purpose of estimating SQA.

Efficient Differential and Differential-Integral Attentive and CNN-based approaches have been explored in this work to classify PPG signal segments as usable or not, further verifying their effectiveness by applying them to four distinct expert-annotated photoplethysmographic datasets. The task they performed is framed as a binary classification problem and plays a vital role in enabling reliable HMA, particularly in wearable devices. By employing NAS-based techniques, we discovered a baseline CNN architecture that is small enough to be deployed on embedded devices, i.e., with resulting

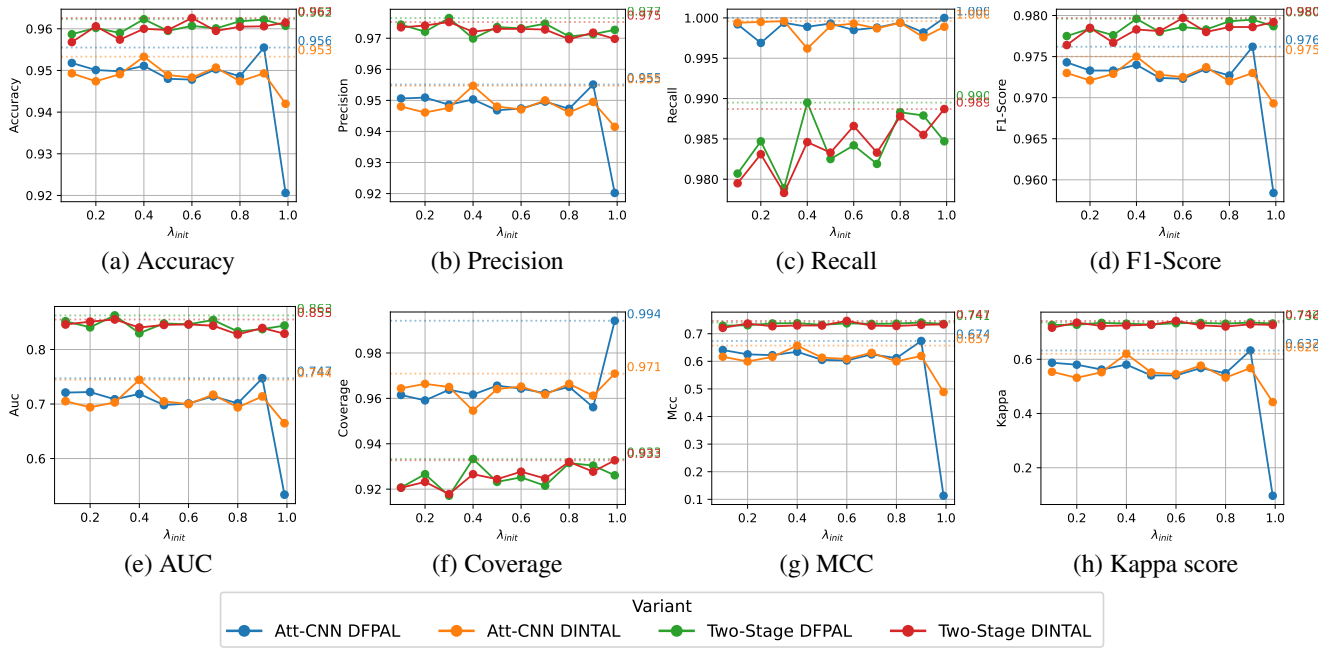


Figure 4. Performance comparison of DFPAL and DINTAL, with and without Two-Stage Classification pipeline, across several values of λ_{init} on GW7 PPG dataset. All the methods were trained for 5 trials and 1000 epochs. The test set performance metrics were then averaged across all trials.

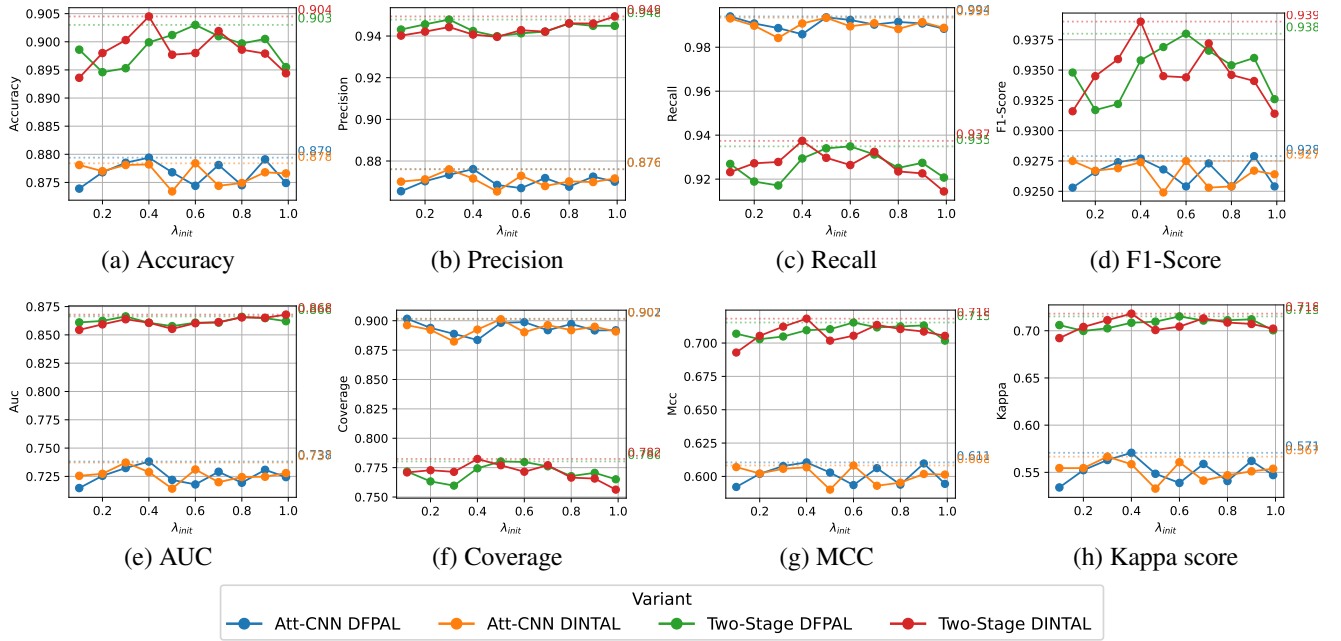


Figure 5. Performance comparison of DFPAL and DINTAL, with and without Two-Stage Classification pipeline, across several values of λ_{init} on RING PPG dataset. All the methods were trained for 5 trials and 1000 epochs. The test set performance metrics were then averaged across all trials.

models requiring less than 500KB available memory, facilitating real-time wearable HMA. Furthermore, the results show that incorporating an additional fine-tuned differential attention layer enhances the performance of the baseline CNN, effectively boosting classification metrics without incurring significant additional computational costs.

One limitation of the present study concerns the fact that all experiments were conducted using sensor data collected from

devices manufactured by a single commercial brand. Although this enabled controlled experimentation and ensured internal consistency in hardware characteristics, firmware behavior, sampling rates, and signal preprocessing pipelines, it may limit the generalization of the findings to devices produced by other manufacturers. Thus, future comparative studies are encouraged.

Production factors, such as sensor properties and architectures, calibration methods, filtering parameters, and signal processing procedures may vary significantly across different brands. As a result, models trained and validated on data from one manufacturer's ecosystem may not directly transfer to devices with different hardware configurations, sampling frequencies, noise characteristics, or embedded preprocessing logic. Due to lack of access to sensor datasets of comparable dimensions from alternative brands, we were unable at this moment to assess cross-brand robustness or quantify potential degradation of performance under heterogeneous hardware environments.

Importantly, sensor-level raw data from competing commercial devices are not usually publicly available. Hence, we restricted our experimental results to data gathered from our protocols, with Samsung devices. Future works would benefit from collaborations between multiple manufacturers, standardized multi-brand benchmarking datasets, as well regulatory agreements that stimulate interoperable and research-accessible multi-manufacturer sensor data. Broader cross-device validation would strengthen external validity and support more comprehensive assessment of model generalization across diverse hardware ecosystems.

References

1. Maniezzo, M., Sandoval, I., Valencio, A. & Lima, R. Improving photoplethysmographic peak detection through higher-order interpolation techniques. In *Proceedings of the 52nd International Conference on Computing in Cardiology (CinC)* (2025).
2. Chatterjee, T., Ghosh, A. & Sarkar, S. Signal quality assessment of photoplethysmogram signals using quantum pattern recognition technique and lightweight cnn module. In *International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 3382–3386 (2022).
3. Gambarotta, N., Aletti, F., Baselli, G. & Ferrario, M. A review of methods for the signal quality assessment to improve reliability of heart rate and blood pressures derived parameters. *Med. & biological engineering & computing* **54**, 1025–1035 (2016).
4. Elgendi, M. Optimal signal quality index for photoplethysmogram signals. *Bioengineering* **3**, 21 (2016).
5. Selvaraj, N., Mendelson, Y., Shelley, K. H., Silverman, D. G. & Chon, K. H. Statistical approach for the detection of motion/noise artifacts in photoplethysmogram. In *International Conference of the IEEE Engineering in Medicine and Biology Society*, 4972–4975 (2011).
6. Li, Q. & Clifford, G. D. Dynamic time warping and machine learning for signal quality assessment of pulsatile signals. *Physiol. measurement* **33**, 1491 (2012).
7. Sivanjaneyulu, Y., Boppu, S. & Manikandan, M. S. Deep fourier magnitude spectrum based signal quality assessment for reducing false alarms under noisy recordings. In *2025 17th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, 1–6, DOI: [10.1109/ECAI65401.2025.11095467](https://doi.org/10.1109/ECAI65401.2025.11095467) (2025).
8. Sun, X., Yang, P. & Zhang, Y.-T. Assessment of photoplethysmogram signal quality using morphology integrated with temporal information approach. In *International Conference of the IEEE Engineering in Medicine and Biology Society*, 3456–3459 (2012).
9. Li, K., Warren, S. & Natarajan, B. Onboard tagging for real-time quality assessment of photoplethysmograms acquired by a wireless reflectance pulse oximeter. *IEEE Transactions on Biomed. Circuits Syst.* **6**, 54–63 (2011).
10. Sukor, J. A., Redmond, S. & Lovell, N. Signal quality measures for pulse oximetry through waveform morphology analysis. *Physiol. measurement* **32**, 369 (2011).
11. Naeini, E. K., Azimi, I., Rahmani, A. M., Liljeberg, P. & Dutt, N. A real-time ppg quality assessment approach for healthcare internet-of-things. *Procedia Comput. Sci.* **151**, 551–558 (2019).
12. Freitas, P. G., De Lima, R. G., Lucafo, G. D. & Penatti, O. A. Assessing the quality of photoplethysmograms via gramian angular fields and vision transformer. In *31st European Signal Processing Conference (EUSIPCO)*, 1035–1039 (2023).
13. Freitas, P. G., Lima, R. G. D., Lucafo, G. D. & Penatti, O. A. B. Photoplethysmogram signal quality assessment via 1d-to-2d projections and vision transformers. In *International Conference on Quality of Multimedia Experience, Ghent, Belgium, June 20-22*, 165–170 (2023).

14. Lucafo, G., Lima, R., Sandoval, I., Albany, L. & Penatti, O. Correlation-boosted ensemble local patterns for photoplethysmographic signal quality classification. *IEEE Signal Process. Lett.* **32**, 61–65, DOI: [10.1109/LSP.2024.3495573](https://doi.org/10.1109/LSP.2024.3495573) (2025).
15. Silva, L. *et al.* Photoplethysmography signal quality assessment using attentive-cnn models. In *Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*, 308–318 (SBC, 2024).
16. Bahdanau, D., Cho, K. & Bengio, Y. Neural machine translation by jointly learning to align and translate. *3rd Int. Conf. on Learn. Represent. ICLR 2015* (2015).
17. Luong, T., Pham, H. & Manning, C. D. Effective approaches to attention-based neural machine translation. *Proc. 2015 Conf. on Empir. Methods Nat. Lang. Process.* 1412–1421, DOI: [10.18653/v1/D15-1166](https://doi.org/10.18653/v1/D15-1166) (2015).
18. Vaswani, A. *et al.* Attention is all you need. In *Advances in Neural Information Processing Systems*, 5998–6008 (2017).
19. Ye, T. *et al.* Differential transformer. In Yue, Y., Garg, A., Peng, N., Sha, F. & Yu, R. (eds.) *International Conference on Learning Representations*, vol. 2025, 144–164 (2025).
20. Cun, Y. L. *et al.* *Handwritten digit recognition with a back-propagation network*, 396–404 (Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1990).
21. Lucafo, G. *et al.* Signal quality assessment of photoplethysmogram signals using hybrid rule- and learning-based models. In *XIX Congresso Brasileiro de Informática em Saúde (CBIS)* (Brazilian Health Informatics Association (SBIS), 2022).
22. Brown, T. *et al.* Language models are few-shot learners. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M. & Lin, H. (eds.) *Advances in Neural Information Processing Systems*, vol. 33, 1877–1901 (Curran Associates, Inc., 2020).
23. Elsken, T., Metzen, J. H. & Hutter, F. Neural architecture search: a survey. *J. Mach. Learn. Res.* **20**, 1997–2017 (2019).
24. Lima, R. G. *et al.* Neural architecture search for tiny detectors of inter-beat intervals. In *European Signal Processing Conference (EUSIPCO)*, 1085–1089 (2023).
25. Laplante, P. A. (ed.) *Comprehensive dictionary of electrical engineering* (CRC Press, 2018).
26. Cang, Y., Liu, Y., Zhang, X., Zhao, E. & Shi, L. Dint transformer (2025).
27. Dalianis, H. Evaluation metrics and evaluation. *Clin. Text Mining: secondary use electronic patient records* 45–53 (2018).
28. Garcia Freitas, P. *et al.* Method for featuring signals using one-dimensional center asymmetric-symmetric local binary patterns (2025). US Patent App. 18/490,935.
29. Hao, J. & Bo, G. A quality assessment system for ppg waveform. In *International Conference on Circuits and Systems (ICCS)*, 170–175, DOI: [10.1109/ICCS52645.2021.9697132](https://doi.org/10.1109/ICCS52645.2021.9697132) (2021).

Funding

This work was funded by Samsung Eletrônica da Amazônia Ltda., under Brazilian Informatics Law 8.248/91. The approval for human research was conceded by “ConseEP Investiga”, Project ID 074154 and CAAE 07327119.8.0000.5599.

Author contributions statement

R.L., P.G.F. and I.S. conceived the experiment, R.L. and I.S. implemented the methodology, R.L., I.S., A.V. and M.M.O. developed the software and visualization tools, I.S., A.V. and M.M. curated the data, I.S., A.V. and M.M. analysed the final results. All authors reviewed the manuscript.

Additional information

Competing interests

All authors declare current or past employment with the manufacturer of the wearable devices adopted during the acquisition protocol of the datasets used in this work (Samsung Electronics).

Data availability

Due to company regulations, the datasets cannot be shared externally. However, reproducibility studies are welcome using the experimental procedure informed in the Methods section.