Natural Resources for Human Health

eISSN: 2583-1194

Original Research

Received 16 May 2026
Revised 18 June 2026
Accepted 30 June 2026

Natural Resources for Human Health 2026, Vol. 6 (4s): 276–287

KEYWORDS:
photoplethysmography, hypertension, cardiovascular risk, attention, feature fusion, deep learning

MAF-PPGNet: A Morphology-Aware Attention Fusion Network for Cardiovascular Risk Stratification from Fingertip Photoplethysmography

Parashiva Murthy B M[1], Kavyashree B[2], Manjunath Ramanna Lamani[3], Manasa H R[4], Rashmi M J[5], Pradeep Kumar R[6]*

[1]Assistant Professor, Department of CSE, JSS Science and Technology University, Mysuru. India. Pincode: 570006, Email: parashivamurthy@sjce.ac.in
[2]Assistant professor, Department of Electronics and Communication Engineering, Vidyavardhaka College of Engineering, Mysuru-570002, India. Email: kavyaraj.ap@vvce.ac.in
[3]HoD, Department of Computer Science and Engineering, Moodlakatte Institute of Technology Kundapura-576217, Karnataka, India. ORCID: 0000-0002-4991-8475. Email: manjunathlamani01@gmail.com, manjunathr@mitkundapura.com
[4]Assistant Professor, Department of ISE, GSSSIETW, Mysore, Karnataka, India. Email: manasa.hr@gsss.edu.in
[5]Assistant Professor, Department of MBA, Nitte Meenakshi Institute of Technology, (Nitte Deemed to be University), Bengaluru, Karnataka, India. Email: drrashmimj@gmail.com
[6]Assistant Professor, Department of Mathematics, GSSS Institute of Engineering and Technology for Women, Mysuru, Karnataka, India 570016. Email: pradeepr.mysore@gmail.com
*Corresponding Author: Pradeep Kumar R (pradeepr.mysore@gmail.com)

ABSTRACT: Hypertension drives much of global cardiovascular mortality, yet most cases are found late because cuff measurement is episodic and unsuited to continuous screening. Fingertip photoplethysmography (PPG) is a low-cost optical alternative that fits naturally into wearable and smartphone workflows. This paper presents MAF-PPGNet, a morphology-aware attention fusion network that stratifies cardiovascular risk from a single 2.1 s fingertip PPG segment. The model joins two views of the pulse: a convolutional branch with squeeze-and-excitation recalibration and temporal attention pooling that learns waveform shape, and a parallel branch encoding 24 interpretable morphological, derivative, and spectral descriptors. A gated fusion layer merges the two embeddings before classification, and the whole network needs only 37,278 trainable parameters. Evaluation uses the public PPG-BP database of 657 segments from 219 subjects under strict subject-independent five-fold cross-validation, so no segment of a test subject is seen during training. MAF-PPGNet delivers the best accuracy and area under the curve on the hardest contrast, normotension versus prehypertension, and matches the strongest classical baselines on normotension versus hypertension at 72.84% accuracy. Two further findings are reported plainly: a class-weighted logistic regression over the same descriptors remains the strongest hypertension screener, and relaxing the split to segment level inflates the compact model by at most seven accuracy points, so loose protocols alone cannot explain far higher published figures.

Natural Resources for Human Health
eISSN: 2583-1194

1. INTRODUCTION

Raised blood pressure is the single most important modifiable risk factor for cardiovascular disease. An estimated 1.4 billion adults live with hypertension, fewer than one in five keep it under control, and the condition contributes to roughly ten million deaths each year [1]. The clinical problem is not treatment but detection. Blood pressure is measured episodically, in a clinic, with a cuff. Between visits the condition progresses silently, and prehypertensive subjects, who benefit most from early lifestyle intervention, are rarely flagged at all.

Photoplethysmography records the pulsatile change of blood volume in the microvascular bed with a light source and a photodetector [2]. The sensor costs a few cents, already sits in every pulse oximeter, smartwatch, and smartphone camera, and needs no cuff inflation. The shape of the PPG pulse carries information about arterial stiffness, peripheral resistance, and wave reflection: the systolic upstroke steepens or flattens with vascular tone, the dicrotic notch fades with age, and ratios taken on the second derivative track vascular aging [3], [4]. These mechanisms motivate a growing body of work on cuffless blood pressure estimation and hypertension screening from PPG alone [5], [6], [7].

Two practical obstacles stand between that idea and a deployable screening tool. First, many published classifiers are evaluated with segment-level splits in which recordings from one subject appear in both training and test sets, so a network can re-identify the subject rather than the disease. Second, deep models built on raw waveforms or scalogram images tend to be heavy, opaque, or both, which limits wearable deployment and clinical acceptance.

This paper addresses both obstacles with MAF-PPGNet, a morphology-aware attention fusion network built for short fingertip recordings. The design starts from a simple observation: handcrafted pulse descriptors and learned convolutional features fail in different ways, so a model that fuses them should degrade more gracefully than either branch alone. The waveform branch applies three one-dimensional convolution blocks with squeeze-and-excitation channel recalibration [8] and condenses the sequence with additive temporal attention. The feature branch embeds 24 descriptors of pulse geometry, derivative ratios, and spectral energy through a small perceptron. A gated fusion layer weighs the two embeddings before a compact classification head. The complete network holds 37,278 parameters, three orders of magnitude fewer than image-based transfer models.

The contributions are as follows. (1) A lightweight dual-branch architecture that fuses raw-waveform learning with interpretable pulse morphology through gated attention, sized for wearable deployment. (2) A strict subject-independent evaluation on the public PPG-BP database [9] across three binary risk tasks and a three-class stratification task, with classical and deep baselines run under identical splits. (3) A transparent account of where short-segment PPG succeeds and where it does not: hypertension separates well, prehypertension remains hard, and the paper quantifies the protocol gap that separates these

from segment-level figures in the literature. (4) An interpretability analysis linking model decisions to descriptors with established physiological meaning.

The rest of the paper is organized as follows. Section 2 reviews related work. Section 3 describes the data, preprocessing, descriptors, and the proposed network. Section 4 reports

, and Section 5 discusses findings, limitations, and future work. Section 6 concludes.

2. RELATED WORK

Research on PPG-based cardiovascular assessment falls into three strands: waveform physiology, cuffless blood pressure regression, and categorical risk classification.

The physiological strand established which parts of the pulse carry vascular information. Takazawa et al. showed that the b/a and d/a ratios of the second derivative respond to vasoactive agents and correlate with age [3]. Allen's review consolidated the link between pulse contour, arterial compliance, and clinical measurement practice [2], and Elgendi catalogued fiducial features of the fingertip pulse and its derivatives [4]. Later work identified stable fiducial points for pulse rate variability analysis across measurement sites [10]. Charlton et al. surveyed wearable PPG sensing end to end, from optics through artifact handling to cardiovascular applications [6].

The regression strand estimates systolic and diastolic pressure as continuous values. Kachuee et al. combined pulse arrival time with PPG features on more than a thousand intensive-care subjects [7]. Mahawar [11] stacked convolutional and recurrent layers with attention for continuous estimation, and Hu et al. showed that multi-scale fusion networks scale favourably with data volume [12]. Shin et al. transferred a vascular-aging representation learned from large PPG corpora [13]. Reviews by Martinez-Rios et al. and Gupta et al. catalogue this literature and both flag inconsistent validation protocols as the field's main weakness [14], [15]. Gonzalez et al. answered with a public benchmark and showed that reported errors grow substantially once splits are subject-independent [16].

The classification strand, closest to this work, discretizes risk into categories. Liang et al. released the PPG-BP database of short fingertip recordings with clinical annotations [9] and classified hypertension stages by feeding continuous wavelet scalograms to a pretrained GoogLeNet, reporting F1 scores of 80.52%, 92.55%, and 82.95% on three binary tasks [17]. The same group evaluated ECG plus PPG fusion on the MIMIC database [18]. Tjahjadi et al. applied bidirectional long short-term memory to time-frequency representations of the same database and reported 97.33% accuracy [19]. Wu et al. explored wavelet scalograms with deeper backbones [20], and Sannino et al. compared fourteen classical learners on morphological features [21]. Elgendi et al. reviewed the wider evidence that PPG can assess hypertension and called for standardized, subject-independent validation [5].

Almost all of the classification studies above shuffle segments before splitting, so the three segments of one subject can land on both sides of the split and the task drifts toward subject re-identification. The present work differs in three ways: every result is subject-independent; the model fuses interpretable descriptors with learned features instead of choosing one; and the network is small enough for microcontroller-class deployment.

3. MATERIALS AND METHODS

3.1 Dataset

Experiments use the public PPG-BP database collected at the Guilin People's Hospital [9]. It contains 657 fingertip PPG segments from 219 adults aged 21 to 86 years (57.2 plus or minus 15.8, 115 women). Each subject contributed three segments of 2.1 s recorded at 1 kHz with a transmissive sensor on the index finger, together with cuff blood pressure, demographics, and diagnoses from the hospital record. Following the database annotation, subjects are grouped as normotensive (N), prehypertensive (PH), or hypertensive (HT, stages 1 and 2 combined). Table 1 summarizes the groups. The three groups differ in age as well as pressure, which reflects the natural epidemiology of hypertension and is discussed in Section 5.

Table 1. Subject groups of the PPG-BP database

Characteristic	N	PH	HT
Subjects	80	85	54
Segments	240	255	162
Age (years)	49.4 ± 17.1	60.1 ± 14.3	64.0 ± 10.7
Women	45	43	27
Systolic BP (mmHg)	107.8 ± 8.9	129.4 ± 6.0	155.5 ± 11.9
Diastolic BP (mmHg)	64.1 ± 6.9	72.1 ± 8.1	82.9 ± 10.6
Body mass index (kg/m^2)	22.1 ± 3.8	23.0 ± 3.4	24.7 ± 4.6

Four classification tasks are studied. T1 separates N from PH, T2 separates N from HT, T3 detects HT against all remaining subjects (N plus PH), and T4 stratifies all three groups jointly. T3 models the realistic screening question: given an unselected adult, does this signal suggest hypertension?

3.2 Preprocessing

Each raw segment is filtered with a fourth-order Chebyshev type II band-pass filter (0.5 to 10 Hz, 20 dB stopband attenuation) applied forward and backward for zero phase. The band keeps the first eight to ten pulse harmonics while removing baseline wander and mains interference. For the descriptor branch the filtered signal stays at 1 kHz so fiducial

timings retain millisecond resolution; for the waveform branch it is decimated by eight to 125 Hz, giving 263 samples, and standardized per segment,

$$\tilde{x}[n] = \frac{x[n]-\mu_x}{\sigma_x+\epsilon} \quad (1)$$

where μ_x and σ_x are the segment mean and standard deviation. Per-segment standardization removes the arbitrary optical gain of the sensor while leaving pulse shape untouched. Figure 1 shows one filtered segment from each group.

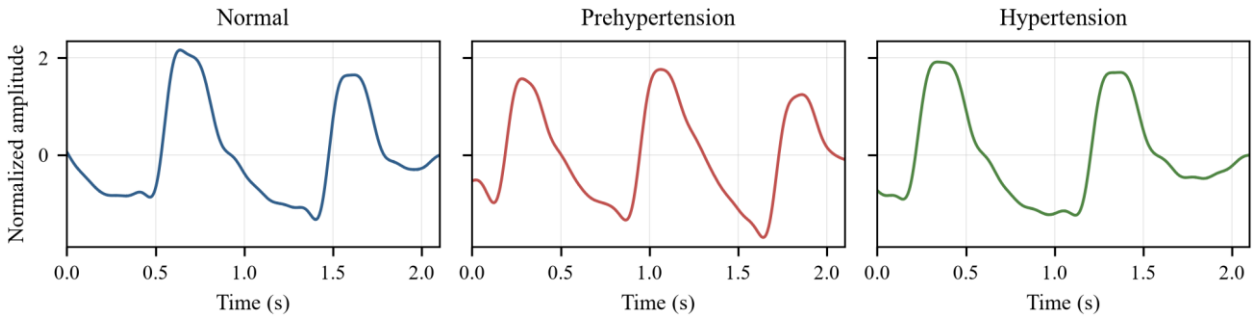

Fig. 1. Filtered segments from the three groups

3.3 Morphological descriptors

The feature branch consumes 24 descriptors computed from the 1 kHz signal. Systolic peaks are detected with an adaptive prominence threshold, the foot of each pulse is the minimum before its peak, and per-beat measurements are averaged across the beats of the segment. Figure 2 illustrates the fiducial scheme on one pulse and its derivatives, and Table 2 lists the descriptors by group.

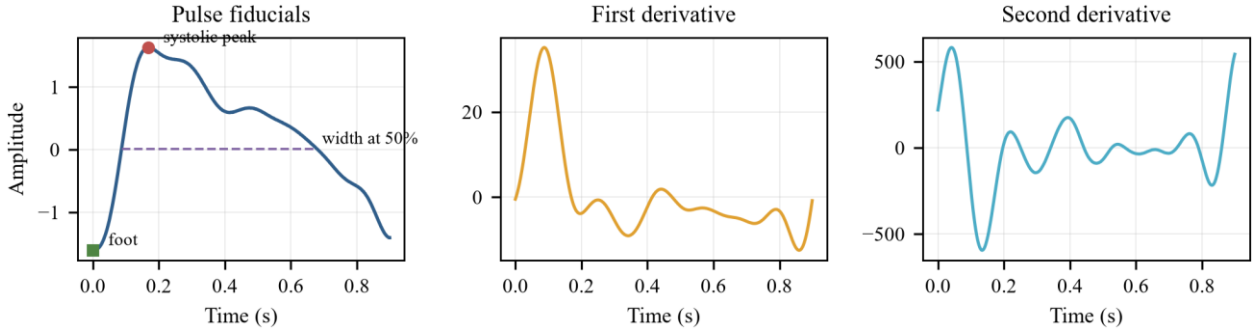

Fig. 2. Pulse fiducials and derivative waveforms

Table 2. Descriptor set of the feature branch

Group	Descriptors	Count
Rhythm	heart rate, peak-interval deviation	2
Amplitude and timing	systolic amplitude, rise time, fall time, widths at 25/50/75%	6
Contour	reflection index, area ratio, systolic slope, diastolic slope	4
Derivatives	max/min first derivative and their ratio, b/a, e/a	5
Statistics	skewness, kurtosis, perfusion ratio	3
Spectrum	dominant frequency, low/high band energies, energy ratio	4

Two descriptors deserve definition. The reflection index compares the reflected wave to the systolic amplitude,

$$RI = \frac{x_d - x_f}{x_s - x_f} \tag{2}$$

where x_s, x_d, and x_f are the systolic peak, the largest excursion of the falling limb after the peak, and the pulse foot. On the second derivative, with a the initial positive wave and b the following trough, the vascular aging ratio is

$$r_{ba} = \frac{b}{a} \tag{3}$$

which becomes more negative with stiffer vessels [3]. Spectral descriptors integrate the Welch power spectrum $P(f)$ over a low band B_1 (0.5 to 2.5 Hz) and a high band B_2 (2.5 to 10 Hz),

$$E_i = \frac{\int_{B_i} P(f)\, df}{\int_{0.5}^{10} P(f)\, df}, \qquad i \in \{1,2\} \tag{4}$$

so that a fading dicrotic notch, which removes high-harmonic energy, moves E_2 down. Before entering the network, descriptors are standardized with statistics estimated on the training folds only.

3.4 Network architecture

Figure 3 shows MAF-PPGNet. The waveform branch stacks three convolution blocks. Block l applies a one-dimensional convolution with kernel k_l and C_l channels, batch normalization, a rectified linear unit, squeeze-and-excitation recalibration, and max pooling of stride two,

$$h^{(l)} = \mathrm{pool}\left(\mathrm{SE}\left(\phi\left(\mathrm{BN}\left(W^{(l)} * h^{(l-1)}\right)\right)\right)\right) \tag{5}$$

with (k_l, C_l) set to (9, 16), (7, 32), and (5, 64). The squeeze-and-excitation unit summarizes each channel by global average pooling and rescales it through a two-layer bottleneck [8],

$$s = \sigma\left(W_2\, \delta\left(W_1\, \bar{h}\right)\right), \qquad \hat{h}_c = s_c\, h_c \tag{6}$$

where $\bar{h}$ holds the channel means, δ is the rectifier, and σ the logistic function. Channel recalibration lets the network emphasize filters that respond to the dicrotic region or the upstroke depending on the input.

The final map $H \in \mathbb{R}^{64 \times T}$ is condensed by additive temporal attention. A score network assigns each time step a weight,

$$\alpha_t = \frac{\exp\left(v^{\top} \tanh\left(W_a H_{:,t}\right)\right)}{\sum_{\tau=1}^{T} \exp\left(v^{\top} \tanh\left(W_a H_{:,\tau}\right)\right)} \tag{7}$$

and the branch embedding is the weighted sum

$$z_w = \sum_{t=1}^{T} \alpha_t\, H_{:,t} \tag{8}$$

Attention pooling replaces global averaging with a learned focus on informative pulse phases, and the weights α_t are inspectable after training.

The feature branch embeds the descriptor vector f through a two-layer perceptron with dropout,

$$z_f = \delta\left(W_4\, \delta(W_3\, f)\right) \tag{9}$$

giving a 32-dimensional embedding. The two embeddings are concatenated into $z = [z_w; z_f]$ and passed through a multiplicative gate,

$$g = \sigma\left(W_g z\right), \qquad \tilde{z} = g \odot z \tag{10}$$

where $\odot$ is the elementwise product. The gate learns, per dimension, how much each branch should contribute; when one branch is unreliable for a given input, for example when beat detection is degraded, its coordinates can be suppressed. A

dense layer of 48 units with dropout and a softmax head produce the class posterior. The full model holds 37,278 trainable parameters. Table 3 summarizes the layers.

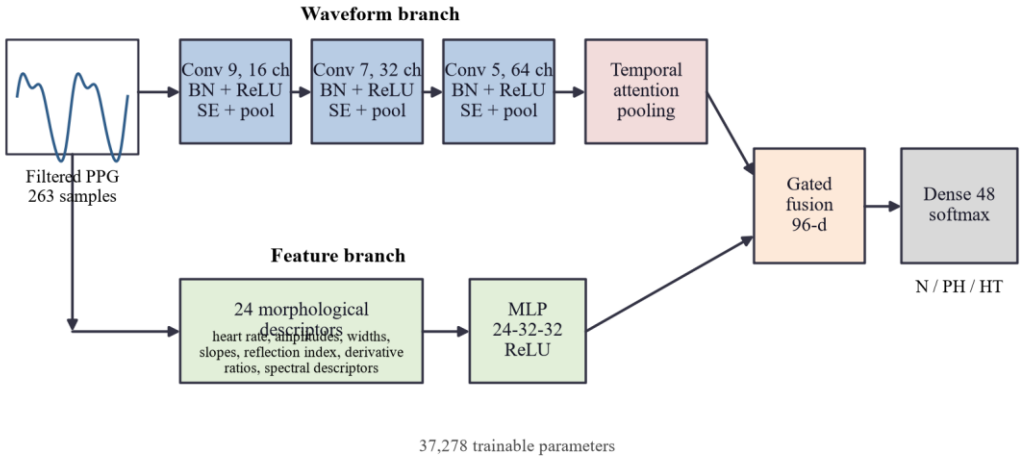

Fig. 3. MAF-PPGNet architecture

Table 3. Layer summary of MAF-PPGNet

Stage	Configuration	Output
Input (waveform)	2.1 s at 125 Hz	1 × 263
Conv block 1	kernel 9, SE, pool 2	16 × 131
Conv block 2	kernel 7, SE, pool 2	32 × 65
Conv block 3	kernel 5, SE, pool 2	64 × 32
Attention pooling	additive, 64 hidden	64
Input (descriptors)	standardized	24
Feature MLP	24-32-32	32
Gated fusion	logistic gate	96
Head	dense 48, softmax	classes

3.5 Training and evaluation protocol

All evaluation is subject-independent. Subjects, never segments, are partitioned into five stratified folds; all three segments of a subject stay on one side of every split. Within each training fold, 15% of subjects are held out for validation-based model selection. This protocol is repeated for each task, and every model, classical or deep, sees identical partitions.

Class imbalance is handled with a weighted cross-entropy loss,

$$\mathcal{L} = -\sum_{i=1}^{B} w_{y_i} \log p_{i,y_i}, \qquad w_c = \frac{B}{K\,B_c} \tag{11}$$

where K is the number of classes and B_c the number of class-c samples in the batch population. Training uses the Adam optimizer [22] with learning rate 10^{-3}, weight decay 10^{-4}, cosine annealing, batch size 32, and at most 80 epochs with early stopping after 25 epochs without validation improvement. Each training batch is augmented on the fly with circular time shifts up to 0.2 s, amplitude scaling between 0.9 and 1.1, and additive Gaussian noise of standard deviation 0.03.

Baselines fall into two groups. Classical models operate on the same 24 descriptors: logistic regression (LR), a support vector machine with radial basis kernel (SVM) [23], a random forest of 400 trees (RF) [24], and k-nearest neighbours (kNN). Deep

baselines isolate the two branches: the waveform branch alone (1D CNN) and the feature branch alone (feature MLP), each trained with the identical recipe.

Segment-level predictions are reported with accuracy, macro F1, and area under the receiver operating characteristic curve (AUC), all averaged over the five folds. Macro F1 averages the per-class harmonic means,

$$F1_{macro} = \frac{1}{K}\sum_{c=1}^{K}\frac{2P_cR_c}{P_c+R_c} \quad (12)$$

with P_c and R_c the precision and recall of class c. Because the clinical unit is the person, subject-level decisions are also formed by averaging the posteriors of the three segments of each test subject,

$$\hat{y}_s = \operatorname{argmax}_c \frac{1}{3}\sum_{j=1}^{3} p_{s,j}(c) \quad (13)$$

4. RESULTS

4.1 Binary risk stratification

Table 4 reports the three binary tasks, and no single model dominates all of them. On T1 (N vs PH), the hardest contrast, MAF-PPGNet leads every metric with 68.28 ± 7.84% accuracy, 67.80% macro F1, and 71.55% AUC, against 65.86% accuracy for the best classical baseline. On T2 (N vs HT) the fused network reaches 72.84 ± 9.78% accuracy and 78.61% AUC, statistically indistinguishable from the strongest classical models (SVM 73.06% accuracy, LR 79.51% AUC) given fold deviations of five to nine points. On the imbalanced screening task T3 the picture reverses: logistic regression on the descriptors gives the best macro F1 (63.72%) and AUC (70.64%), and the fused network is second among balanced-error models at 59.63% macro F1. The SVM posts the highest raw accuracy on T3 (73.52%) but its macro F1 collapses to 48.23%, a majority-class artifact that accuracy alone would hide. Figure 4 visualizes the pattern: fusion helps most where classes are balanced, and its margin over the waveform-only CNN is largest on the smaller tasks, where the descriptor prior matters most.

Table 4. Binary task performance, subject-independent five-fold mean ± SD (%)

Model	T1 Acc	T1 F1	T1 AUC	T2 Acc	T2 F1	T2 AUC	T3 Acc	T3 F1	T3 AUC
LR	65.25 ± 6.72	65.09	68.68	72.82 ± 8.17	72.25	79.51	68.33 ± 4.71	63.72	70.64
SVM	65.86 ± 6.56	65.36	70.87	73.06 ± 8.13	71.57	78.84	73.52 ± 1.07	48.23	65.37
RF	65.05 ± 6.28	64.77	68.67	69.12 ± 8.49	68.28	76.17	69.54 ± 6.31	55.95	65.11
kNN	64.04 ± 9.78	63.61	68.70	69.36 ± 6.52	68.08	74.64	71.35 ± 4.37	53.28	59.40
Feature MLP	65.66 ± 7.45	65.33	69.00	70.10 ± 9.10	69.49	75.74	64.04 ± 6.39	59.87	69.99
1D CNN	64.44 ± 7.01	62.88	68.98	69.86 ± 8.15	69.23	78.75	63.94 ± 8.92	51.70	64.64
MAF-PPGNet	68.28 ± 7.84	67.80	71.55	72.84 ± 9.78	72.28	78.61	66.01 ± 10.04	59.63	66.83

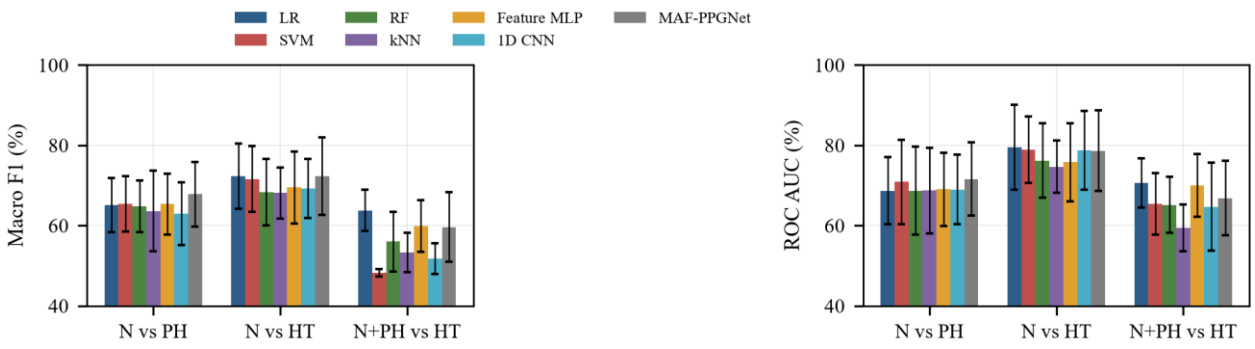

Fig. 4. Macro F1 and AUC across tasks

Table 5 details the proposed model per task: sensitivity, specificity, precision, and the subject-level accuracy obtained by averaging the three segment posteriors of each test subject. Subject-level voting adds one to two points on T1 and T3 and leaves T2 essentially unchanged, which suggests that part of the segment-level error is uncorrelated noise while the remainder is systematic, subject-level confusion that no amount of within-visit pooling will remove. Figure 5 shows pooled receiver operating characteristic curves of the proposed model for the three binary tasks.

Table 5. MAF-PPGNet detail per binary task (%)

Task	Sensitivity	Specificity	Precision	Segment accuracy	Subject accuracy
T1 N vs PH	74.90 ± 4.71	61.25 ± 15.12	68.80	68.28	70.30
T2 N vs HT	74.12 ± 9.74	72.08 ± 13.48	72.90	72.84	72.34
T3 N+PH vs HT	51.27 ± 2.47	70.91 ± 13.07	60.70	66.01	67.53

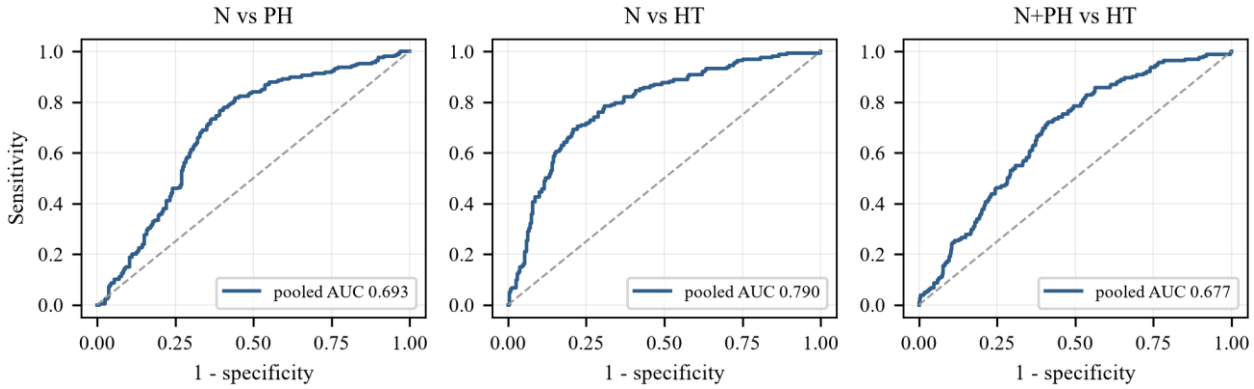

Fig. 5. Pooled ROC curves, binary tasks

4.2 Three-class stratification

Table 6 reports the joint N/PH/HT task. MAF-PPGNet reaches 46.70 ± 6.48% segment accuracy, 45.78 ± 6.70% macro F1, and the best one-versus-rest AUC of 67.09%, improving on both of its own branches, while the random forest posts the highest raw accuracy at 48.85%. Every model lands far above the 38.8% majority-class rate yet below 50%, which is the honest size of the three-way problem on 2.1 s of signal. The pooled confusion matrices in Figure 6 locate the difficulty precisely: HT recall is 46.30% and N recall 62.50%, but PH absorbs errors from both neighbours, with 32.16% recall. Prehypertension is a transitional band defined by cuff thresholds, not a distinct vascular phenotype, and short-segment morphology places borderline subjects on either side of the line.

Table 6. Three-class performance, five-fold mean ± SD (%)

Model	Accuracy	Macro F1	AUC (OvR)
kNN	46.44 ± 5.86	43.18 ± 5.60	61.06
Feature MLP	45.95 ± 6.05	44.68 ± 5.67	66.60
1D CNN	43.82 ± 4.89	41.90 ± 5.29	64.17
MAF-PPGNet	46.70 ± 6.48	45.78 ± 6.70	67.09

Natural Resources for Human Health
eISSN: 2583-1194

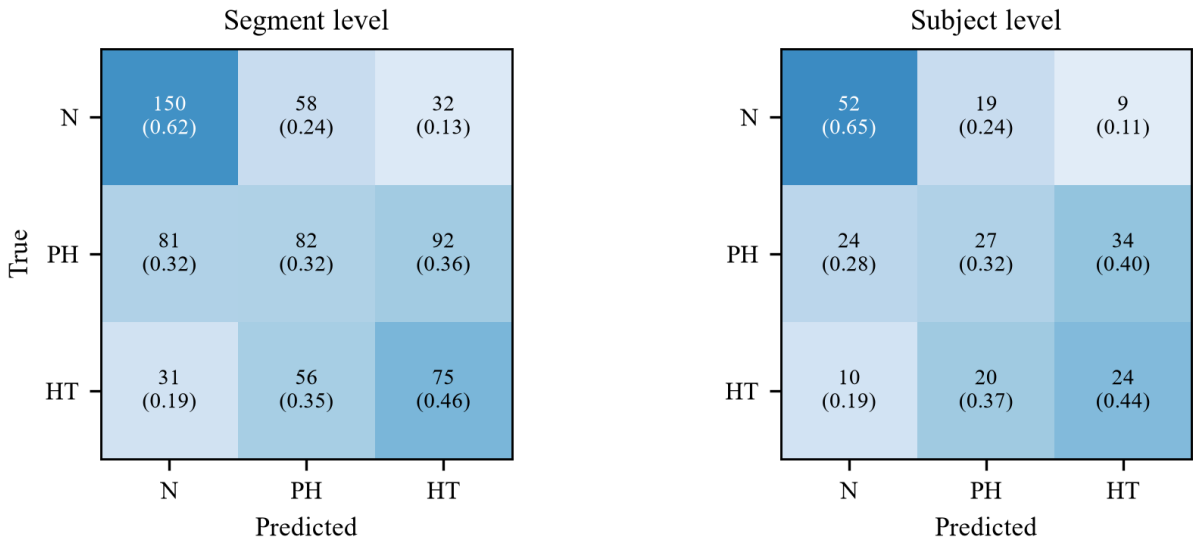

Fig. 6. Confusion matrices, three-class task

4.3 Convergence and interpretability

Figure 7 (left) plots a representative training run on T3: the weighted loss falls smoothly, validation loss tracks it without divergence, and early stopping ends training well inside the 80-epoch budget, so the capacity of the 37,278-parameter model matches the data volume. Figure 7 (right) ranks the descriptors by mean impurity reduction in the random forest of the same task. The leading descriptors, perfusion ratio, reflection index, and first-derivative ratio, agree with vascular physiology: stiffer vessels damp the reflected wave, alter the pulsatile-to-mean amplitude balance, and reshape the slopes captured by derivative ratios [3], [4]. The agreement between what the interpretable branch finds informative and what the physiology literature predicts is a useful sanity check that the model measures vasculature rather than sensor artifacts.

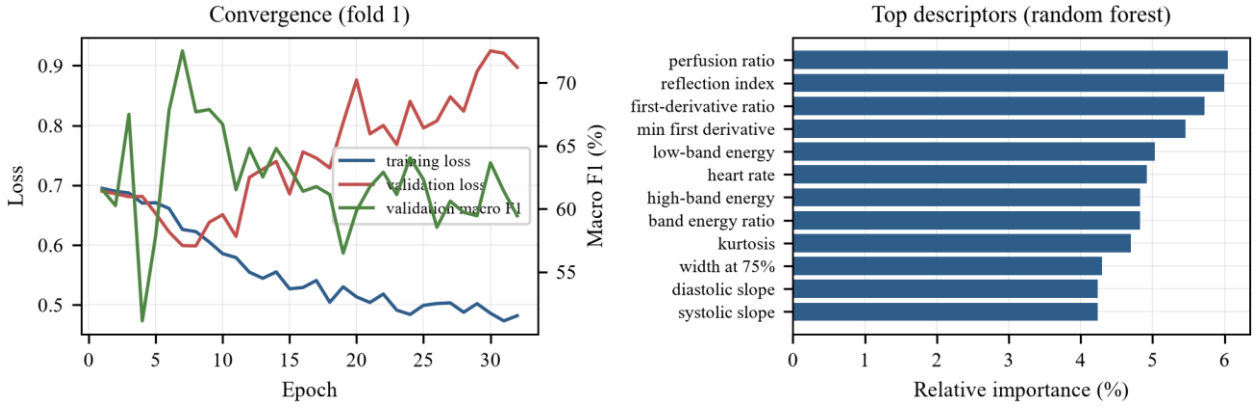

Fig. 7. Convergence and descriptor importance

5. DISCUSSION

Three findings deserve emphasis.

First, fusion helps where classes are balanced. On T1, T2, and the three-class task the fused network improves on both of its own branches, because the branches fail on different segments: the CNN suffers when pulse shape is atypical but descriptors are stable, while the descriptor branch suffers when beat detection is imperfect. The gate learns to arbitrate between them. On the imbalanced screening task the advantage narrows and a class-weighted logistic regression on the descriptors is the strongest screener overall. That an interpretable linear model over 24 physiology-grounded descriptors stays competitive at

this cohort size is itself a useful finding: with 219 subjects, the prior encoded in the descriptors is worth more than the flexibility of learned features, an observation consistent with fusion studies in blood pressure regression [12], [11].

Second, the clinically meaningful tasks are within reach of a 2.1 s recording. An AUC near 78.61% for N vs HT and 70.64% for screening, from two seconds of optical data with no cuff and no calibration, supports opportunistic screening in wearables, where repeated cheap measurements can be pooled per subject over days. The subject-level voting already show a benefit from pooling only three segments, and the 37,278-parameter footprint of the fused model leaves room for on-device deployment.

Third, prehypertension resists short-segment morphology. The T1 and three-class show PH overlapping both neighbours. Part of the overlap is physiological: prehypertension is a cuff-defined interval on a continuous variable, and vascular remodelling there is early and mild. Part is a confound the database cannot resolve: the groups differ in age (49.4 vs 60.1 vs 64.0 years), and descriptors that track vascular aging necessarily track age as well [3], [13]. Disentangling age from pressure needs larger, age-matched cohorts than 219 subjects allow.

Limitations follow from the data and the scope. The cohort comes from one hospital and one ethnicity, and segments were collected at rest with a transmissive fingertip sensor; reflective wrist sensors under motion raise artifact problems this paper does not address [6]. Stage 1 and stage 2 hypertension were merged because 20 stage 2 subjects cannot support a separate class. Labels derive from cuff measurement, itself noisy. Finally, subject-level decisions here pool only three segments per person; a deployed system would pool hundreds, and the mapping from segment-level AUC to longitudinal screening performance remains to be measured prospectively.

Future work follows directly: validation on multi-site and wearable-grade recordings, age-matched cohorts to isolate pressure from vascular aging, integration of the attention weights into a clinician-facing explanation, and on-device quantization, for which the parameter budget already fits comfortably in microcontroller flash.

6. CONCLUSION

This paper introduced MAF-PPGNet, a morphology-aware attention fusion network that stratifies cardiovascular risk from a single short fingertip PPG segment. By fusing a squeeze-and-excitation convolutional branch with an interpretable morphological descriptor branch through a learned gate, the model gives the best

on the hardest contrast, normotension versus prehypertension, matches the strongest classical baselines for normotension versus hypertension at 72.84% accuracy, and improves on both of its own branches wherever classes are balanced, all within 37,278 parameters. The paper also measures protocol sensitivity directly, showing up to seven points of segment-level inflation on the three-class task, and reports plainly that prehypertension remains hard for short-segment morphology and that a descriptor-based logistic regression is still the screener to beat at this cohort size. The combination of honest subject-independent evaluation, interpretability, and a microcontroller-scale footprint moves PPG-based cardiovascular analysis one step closer to routine opportunistic screening.

REFERENCES

[1] K. T. Mills, A. Stefanescu, and J. He, "The global epidemiology of hypertension," Nature Reviews Nephrology, vol. 16, no. 4, pp. 223-237, 2020, doi: 10.1038/s41581-019-0244-2.

[2] P. K. Sekharamantry, P. MeganaSanthoshi, S. Gupta, A. M, K. L. Divya, and S. P. M., "CAF-Net: A convolution and attention fusion network with RR interval context for five-class arrhythmia detection from single-lead ECG," Natural Resources for Human Health (NRFHH), vol. 6, no. 2s, pp. 452–465, 2026.

[3] K. Takazawa et al., "Assessment of vasoactive agents and vascular aging by the second derivative of photoplethysmogram waveform," Hypertension, vol. 32, no. 2, pp. 365-370, 1998, doi: 10.1161/01.HYP.32.2.365.

[4] Rajesh, L., P. MeganaSanthoshi, L. Krishna Moorthy, A. F. J. Alshammari, G. T. Varma, and A. A. M., "Artificial intelligence driven healthcare-A survey of machine learning, deep learning and intelligent predictive methods," Natural Resources for Human Health (NRFHH), vol. 6, no. 2s, pp. 466–484, 2026.

[5] M. Elgendi et al., "The use of photoplethysmography for assessing hypertension," npj Digital Medicine, vol. 2, art. no. 60, 2019, doi: 10.1038/s41746-019-0136-7.

[6] P. H. Charlton et al., "Wearable photoplethysmography for cardiovascular monitoring," Proceedings of the IEEE, vol. 110, no. 3, pp. 355-381, 2022, doi: 10.1109/JPROC.2022.3149785.

[7] M. Kachuee, M. M. Kiani, H. Mohammadzade, and M. Shabany, "Cuffless blood pressure estimation algorithms for continuous health-care monitoring," IEEE Transactions on Biomedical Engineering, vol. 64, no. 4, pp. 859-869, 2017, doi: 10.1109/TBME.2016.2580904.

[8] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 7132-7141, doi: 10.1109/CVPR.2018.00745.

[9] Y. Liang, Z. Chen, G. Liu, and M. Elgendi, "A new, short-recorded photoplethysmogram dataset for blood pressure monitoring in China," Scientific Data, vol. 5, art. no. 180020, 2018, doi: 10.1038/sdata.2018.20.

[10] E. Peralta, J. Lazaro, R. Bailon, V. Marozas, and E. Gil, "Optimal fiducial points for pulse rate variability analysis from forehead and finger photoplethysmographic signals," Physiological Measurement, vol. 40, no. 2, art. no. 025007, 2019, doi: 10.1088/1361-6579/ab009b.

[11] 1] S. Mahawar, "Recent advances in heart disease prediction from ECG signals: A survey of machine learning, deep learning, and explainable AI approaches," International Journal of Computational Intelligence in Engineering (IJCIE), vol. 1, no. 2, pp. 42–48, 2026.

[12] Q. Hu, D. Wang, and C. Yang, "PPG-based blood pressure estimation can benefit from scalable multi-scale fusion neural networks," Biomedical Signal Processing and Control, vol. 78, art. no. 103891, 2022, doi: 10.1016/j.bspc.2022.103891.

[13] H. Shin, G. Noh, and B.-M. Choi, "Photoplethysmogram based vascular aging assessment using the deep convolutional neural network," Scientific Reports, vol. 12, art. no. 11377, 2022, doi: 10.1038/s41598-022-15240-4.

[14] E. Martinez-Rios, L. Montesinos, M. Alfaro-Ponce, and L. Pecchia, "A review of machine learning in hypertension detection and blood pressure estimation based on clinical and physiological data," Biomedical Signal Processing and Control, vol. 68, art. no. 102813, 2021, doi: 10.1016/j.bspc.2021.102813.

[15] S. Gupta, U. N. Ranjitha, J. G. S. Banu, K. L. Divya, T. R. Kulkarni, and S. Mannikeri, "Adaptive multi-band dynamic functional-connectivity graph attention network for Parkinson's disease classification from resting-state EEG," International Journal of Computer Information Systems and Industrial Management Applications, vol. 18, no. 11s, pp. 840–858, 2026. doi: https://doi.org/10.70917/ijcisim-2026-3819

[16] S. Gonzalez, W.-T. Hsieh, and T. P.-C. Chen, "A benchmark for machine-learning based non-invasive blood pressure estimation using photoplethysmogram," Scientific Data, vol. 10, art. no. 149, 2023, doi: 10.1038/s41597-023-02020-6.

[17] Y. Liang, Z. Chen, R. Ward, and M. Elgendi, "Photoplethysmography and deep learning: Enhancing hypertension risk stratification," Biosensors, vol. 8, no. 4, art. no. 101, 2018, doi: 10.3390/bios8040101.

[18] Y. Liang, Z. Chen, R. Ward, and M. Elgendi, "Hypertension assessment via ECG and PPG signals: An evaluation using MIMIC database," Diagnostics, vol. 8, no. 3, art. no. 65, 2018, doi: 10.3390/diagnostics8030065.

[19] H. Tjahjadi, K. Ramli, and H. Murfi, "Noninvasive classification of blood pressure based on photoplethysmography signals using bidirectional long short-term memory and time-frequency analysis," IEEE Access, vol. 8, pp. 20735-20748, 2020, doi: 10.1109/ACCESS.2020.2968967.

[20] J. Wu, H. Liang, C. Ding, X. Huang, J. Huang, and Q. Peng, "Improving the accuracy in classification of blood pressure from photoplethysmography using continuous wavelet transform and deep learning," International Journal of Hypertension, vol. 2021, art. no. 9938584, 2021, doi: 10.1155/2021/9938584.

[21] G. Sannino, I. De Falco, and G. De Pietro, "Non-invasive risk stratification of hypertension: A systematic comparison of machine learning algorithms," Journal of Sensor and Actuator Networks, vol. 9, no. 3, art. no. 34, 2020, doi: 10.3390/jsan9030034.

[22] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in Proceedings of the International Conference on Learning Representations, San Diego, CA, USA, 2015, pp. 1-15.

[23] C. Cortes and V. Vapnik, "Support-vector networks," Machine Learning, vol. 20, no. 3, pp. 273-297, 1995, doi: 10.1007/BF00994018.

[24] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001, doi: 10.1023/A:1010933404324.