

Non-Contact Heart Rate Estimation for Moving Subjects via aDual-Trajectory Hybrid Deep Framework

Jianan Tong

Jun Guo

guojun@nwu.edu.cn

Na Peng

Research Article

Keywords: Millimeter-wave radar, contact-free heart-rate monitoring, dynamic scenarios, motion artifact suppression, deep learning

Posted Date: October 22nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9424082/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Non-Contact Heart Rate Estimation for Moving Subjects via a Dual-Trajectory Hybrid Deep Framework

Jianan Tong¹, Jun Guo^{1*}, Na Peng¹

^{1*}College of Computer Science, Northwest University, Xi'an, 710127, Shaanxi Province, China.

*Corresponding author(s). E-mail(s): guojun@nwu.edu.cn;

Contributing authors: 202421965@stumail.nwu.edu.cn; 202421890@stumail.nwu.edu.cn;

Abstract

Contact-free heart-rate monitoring with millimeter-wave Frequency-Modulated Continuous-Wave (FMCW) radar is promising for wearable-free healthcare, but accurate estimation remains difficult when subjects are moving. In dynamic scenarios, gross body motion causes range migration, phase distortion, and Doppler-related interference that can overwhelm the much weaker cardiogenic component. To address this problem, we propose a hybrid *locate-then-reconstruct* framework for robust heart-rate monitoring. In the first stage, a dual-trajectory fusion module with a customized Kalman filter determines a reliable phase-extraction path and retrieves an unquantized raw phase sequence from the original radar measurements. In the second stage, the recovered signal is processed by a Variational Mode Decomposition-enhanced Convolutional Neural Network–Bidirectional Long Short-Term Memory–Autoformer (VCLA) model to suppress residual motion artifacts and reconstruct a cleaner cardiogenic waveform. Template matching is then applied for final heart-rate estimation. Experiments in both single-person and multi-person scenarios show that the proposed method achieves a mean absolute error below 4 beats per minute and consistently outperforms representative baselines.

Keywords: Millimeter-wave radar, contact-free heart-rate monitoring, dynamic scenarios, motion artifact suppression, deep learning

1 Introduction

Millimeter-wave Frequency-Modulated Continuous-Wave (FMCW) radar has emerged as a promising technique for contact-free heart-rate monitoring because it can capture subtle chest micro-motion without physical attachment, making it suitable for long-term sensing in clinical, home-care, and assisted-living settings [1, 2]. In conventional radar-based vital-sign monitoring, heart rate is inferred from periodic phase variations of the reflected signal [3, 4], which is effective when the subject remains quasi-static

[5]. However, in ambulatory or other dynamic scenarios, gross body motion introduces severe range migration, phase distortion, and Doppler-related interference, which corrupt the already weak cardiogenic component and make conventional heart-rate estimation unstable [6].

Existing approaches can be broadly categorized into model-driven and data-driven methods. Model-driven methods, such as Fast Fourier Transform (FFT), Continuous Wavelet Transform (CWT), Empirical Mode Decomposition (EMD)[7], and Variational Mode Decomposition (VMD), provide clear physical interpretability,

but their performance often depends on assumptions of stationarity, separability, or carefully tuned parameters [8–10]. Under strong motion, these methods may suffer from spectral smearing, mode mixing, and unstable component separation. In contrast, deep learning methods can learn motion-tolerant representations directly from data, yet purely data-driven models may overlook physical priors and produce physiologically implausible outputs [11]. Therefore, a practical solution for dynamic heart-rate monitoring should combine motion-aware physical modeling with the representation power of deep neural networks.

To address this challenge, we propose a hybrid *locate-then-reconstruct* framework for dynamic heart-rate monitoring, as illustrated in Fig. 1. In Stage I, a motion-aware tracking module performs dual-trajectory fusion with a customized Kalman filter to determine a reliable phase-extraction path in the range–time domain. This path is then used to retrieve and concatenate the corresponding unquantized raw phase samples from the original radar measurements, yielding a motion-compensated raw phase sequence. In Stage II, the recovered phase signal is fed into a hybrid deep network, termed the VCLA model, which integrates Variational Mode Decomposition (VMD), a Convolutional Neural Network (CNN), a Bidirectional Long Short-Term Memory (BiLSTM) network, and an Autoformer encoder to suppress residual motion artifacts and enhance the weak cardiogenic component. Finally, a template-matching strategy is applied to the reconstructed waveform to refine cardiac cycle estimation and obtain a more stable heart-rate measurement.

The main contributions of this paper are as follows.

- We propose a *locate-then-reconstruct* framework for dynamic heart-rate monitoring with FMCW radar, which decouples motion-aware phase-path localization from cardiogenic waveform recovery.
- We develop a dual-trajectory fusion method for adaptive phase-path localization and a VCLA model for motion-robust cardiogenic waveform reconstruction from unquantized raw phase observations.
- Extensive experiments in both single-person and multi-person scenarios demonstrate that

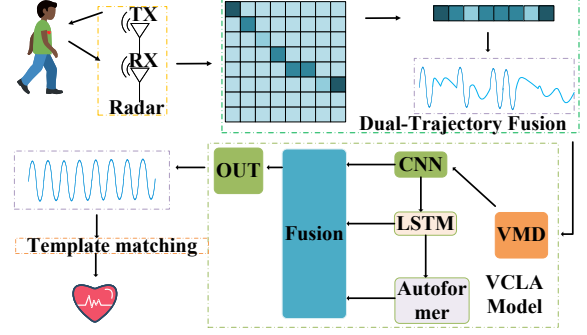


Fig. 1 Overview of the proposed locate-then-reconstruct framework for dynamic heart-rate monitoring. Stage I performs motion-aware phase-path localization and raw phase extraction, Stage II reconstructs the cardiogenic waveform using the VCLA model, and template matching is finally applied for robust heart-rate estimation.

the proposed method achieves a mean absolute error below 4 BPM and consistently outperforms representative baseline methods in both accuracy and robustness.

2 Related Work

2.1 Model-Driven Methods

Model-driven methods estimate heart rate from radar phase signals using physically interpretable processing pipelines. A common strategy is to select a single range bin or dominant trajectory and then perform phase demodulation and spectral analysis [12, 13]. These methods are computationally efficient and effective in quasi-static scenarios, but they are vulnerable to abrupt motion, clutter, and trajectory loss because the selected path can be easily corrupted. To improve robustness, researchers have introduced joint spatiotemporal representations and trajectory-tracking schemes, including range–Doppler or range–speed analysis, phase continuity tracking, and adaptive filtering [14, 15]. In parallel, decomposition methods such as VMD, Empirical Wavelet Transform (EWT), and related variants have been employed to separate respiration, heartbeat, and motion artifacts from the extracted signal [16–18]. However, these methods still largely follow a single-path selection paradigm and often suffer from ambiguous mode attribution when spectral overlap is severe.

2.2 Learning-Based and Hybrid Methods

Deep learning has recently attracted increasing attention for radar-based heart-rate estimation because of its strong capability in nonlinear modeling and feature extraction. CNNs can capture local time-frequency patterns and suppress structured noise, while recurrent models such as LSTM and BiLSTM networks are effective in modeling temporal dependency across cardiac cycles [19, 20]. More recently, Transformer-based architectures and Autoformer variants have shown promise in modeling long-range periodic structure under complex interference [21]. Nevertheless, purely data-driven models may generalize poorly in dynamic environments because they lack explicit physical constraints and may produce physiologically implausible outputs.

To overcome this limitation, hybrid methods that combine signal-processing priors with neural representation learning have begun to emerge [22, 23]. These approaches suggest that physically grounded preprocessing can provide more reliable inputs for subsequent learning modules. Our method follows this direction but differs in two aspects: first, we use dual-trajectory fusion to construct a more robust phase-extraction path instead of relying on a single dominant trajectory; second, we feed the downstream network with the unquantized raw phase sequence retrieved from the preserved original phase matrix, rather than with trajectory coordinates or directly selected bin signals. This design provides a clearer separation between path localization and waveform reconstruction, and yields a more robust basis for dynamic heart-rate estimation.

3 Methodology

The proposed framework follows a locate-then-reconstruct strategy for non-contact heart-rate monitoring under subject motion. Instead of estimating heart rate directly from a fixed range bin, we first localize a reliable phase-extraction path on the range-time map and retrieve the corresponding raw phase sequence from the preserved phase matrix. The extracted sequence is then fed into a waveform reconstruction network to recover the cardiogenic component for subsequent heart-rate estimation.

3.1 Motion-Aware Raw Phase Extraction

For an FMCW radar, the phase variation of the intermediate-frequency signal reflects target displacement and can be expressed as [24]

$$\Delta\phi(t) = \frac{4\pi}{\lambda} [\Delta d(t) + \Delta x(t)], \quad (1)$$

where $\Delta d(t)$ and $\Delta x(t)$ denote gross body motion and subtle chest displacement, respectively. In dynamic scenarios, gross motion dominates the phase variation and causes severe range-bin migration, making direct phase extraction from a fixed bin unreliable.

To address this issue, we estimate a time-varying phase-extraction path from the range-time map. After normalization, contrast enhancement, and median filtering, a coarse trajectory is obtained by multi-threshold weighted centroid detection. Two complementary trajectories are then generated, namely a smoothed interpolation trajectory $\mathbf{t}_A$ and a Kalman-refined trajectory $\mathbf{t}_B$, and fused by local inverse-variance weighting:

$$w_A(i) = \frac{1/\sigma_A^2(i)}{1/\sigma_A^2(i) + 1/\sigma_B^2(i)}, \quad w_B(i) = 1 - w_A(i), \quad (2)$$

$$t_F(i) = w_A(i)t_A(i) + w_B(i)t_B(i), \quad (3)$$

where $\sigma_A(i)$ and $\sigma_B(i)$ denote the local standard deviations of the two trajectories. The fused trajectory $\mathbf{t}_F$ is used only as an indexing path.

Rather than using trajectory coordinates as network input, the preserved phase matrix $\mathbf{P}$ is sampled along $\mathbf{t}_F$ to form a continuous raw phase sequence:

$$\mathbf{p}_{\text{raw}} = [P(1, t_F(1)), P(2, t_F(2)), \dots, P(N, t_F(N))] . \quad (4)$$

In this way, Stage I determines where to read the phase, while Stage II focuses on recovering the underlying cardiogenic waveform from the retrieved raw signal.

3.2 VCLA-Based Waveform Reconstruction

Although Stage I improves phase continuity, the extracted raw phase sequence still contains strong motion artifacts and environmental noise. To

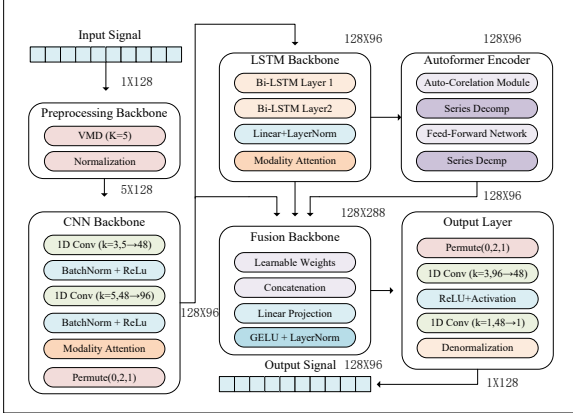


Fig. 2 Architecture of the proposed VCLA network. A raw phase segment of size 1×128 is first decomposed by VMD into $K = 5$ intrinsic mode functions. The resulting multi-channel representation is then processed by a CNN–BiLSTM–Autoformer backbone with dynamic fusion, and the reconstructed waveform is finally fed into the existing template-matching method for heart-rate estimation.

recover the weak cardiogenic component, we formulate Stage II as a supervised waveform reconstruction problem and implement a VMD–CNN–BiLSTM–Autoformer network, termed VCLA, as shown in Fig. 2. The network follows a progressive feature learning strategy: VMD first provides a structured multi-channel representation, CNN extracts local morphology, BiLSTM models temporal dependency, Autoformer enhances long-range periodicity, and a dynamic fusion module aggregates these complementary representations for waveform reconstruction. This design is consistent with the actual implementation in the training code.

Given an input segment $\mathbf{p}_{\text{raw}} \in \mathbb{R}^{1 \times N}$ with $N = 128$, variational mode decomposition (VMD) is first applied to obtain a structured multi-channel representation:

$$\mathbf{X}_{\text{VMD}} = \text{VMD}(\mathbf{p}_{\text{raw}}) \in \mathbb{R}^{K \times N}, \quad (5)$$

where $K = 5$ is the number of intrinsic modes. In the implementation, the VMD parameters are set to $\alpha = 2000$ and $\text{tol} = 10^{-7}$. This preprocessing step separates oscillatory components with different frequency characteristics and reduces the difficulty of directly learning from severely corrupted raw phase observations.

The decomposed signal is then passed through three sequential feature extractors. First, the

CNN branch consists of two one-dimensional convolutional layers with kernel sizes 3 and 5, each followed by batch normalization and ReLU activation. A CNN-specific attention gate is further introduced to emphasize informative local responses. This branch mainly captures local morphological details such as short-term peaks, slopes, and waveform distortions.

Next, the CNN features are fed into a two-layer bidirectional long short-term memory (BiLSTM) network. The bidirectional outputs are projected back to the hidden feature space and normalized by layer normalization, followed by an LSTM-specific attention module. This stage enhances temporal consistency and captures dependencies across adjacent cardiac cycles, which are difficult to model using local convolution alone.

The BiLSTM features are subsequently processed by an Autoformer encoder. In the current implementation, the Autoformer branch uses one encoder layer, two attention heads, a feed-forward dimension of 128, and a moving-average window of 5. Its autocorrelation-based mechanism strengthens periodic representation and complements the local and sequential features extracted by the CNN and BiLSTM branches.

Let $\mathbf{F}_{\text{cnn}}$, $\mathbf{F}_{\text{lstn}}$, and $\mathbf{F}_{\text{auto}}$ denote the features produced by the three stages. Instead of directly averaging them, the network adopts a dynamic fusion module with learnable modal weights:

$$\mathbf{F}_{\text{cat}} = [\theta_1 \mathbf{F}_{\text{cnn}}; \theta_2 \mathbf{F}_{\text{lstn}}; \theta_3 \mathbf{F}_{\text{auto}}], \quad (6)$$

$$\mathbf{F}_{\text{fuse}} = \Phi(\mathbf{F}_{\text{cat}}), \quad (7)$$

where θ_1 , θ_2 , and θ_3 are learnable fusion coefficients, and $\Phi(\cdot)$ denotes a linear projection followed by GELU activation and layer normalization. This design allows the model to adaptively balance local morphology, temporal dependency, and periodic structure according to the corruption level of each input segment.

Finally, the fused feature is decoded into a denoised waveform by a lightweight convolutional head:

$$\hat{\mathbf{y}} = \mathcal{D}(\mathbf{F}_{\text{fuse}}) \in \mathbb{R}^{1 \times N}, \quad (8)$$

where the decoder consists of a 3-point convolution, a ReLU activation, and a 1×1 convolution for signal reconstruction. The network is optimized

Table 1 Hardware setup used in data acquisition and model development.

Category	Item	Specification
Data	Radar transceiver	Skaryilil 77 GHz 4T4R MIMO (RC7711B-AIP)
	Microcontroller	STM32-G473CCU6
	Host computer	Standard laptop for data streaming
Training	Workstation	Standard desktop computer
Reference	Heart-rate monitor	Huawei smart bracelet

using the mean squared error loss

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N} \sum_{t=1}^N (\hat{y}_t - y_t)^2, \quad (9)$$

where $\mathbf{y}$ is the clean reference waveform. In the code, the model is trained with Adam, a learning rate of 5×10^{-4} , weight decay of 5×10^{-5} , and a ReduceLROnPlateau scheduler.

The output of the VCLA model is the reconstructed cardiogenic waveform rather than the final heart-rate value itself. To obtain the final heart rate, the denoised waveform is further fed into an existing template-matching method rather than a separately designed regression head [25].

4 Experiments

4.1 Dataset

Existing dynamic radar datasets are mainly designed for target detection or classification rather than heart rate monitoring. To obtain training data for dynamic heart rate estimation, we adopt a data synthesis strategy. Clean static phase signals are treated as ground truth, and motion interference is introduced according to

$$\Delta\phi_d(t) = \frac{4\pi}{\lambda} \Delta d(t), \quad (10)$$

where $\Delta d(t)$ denotes the motion-induced distance variation. The disturbed phase is then added to $\phi_{\text{clean}}(t)$ to generate synthetic dynamic phase signals. The hardware configuration is summarized in Table 1, and the radar parameters are listed in Table 2.

Eight participants were recruited with written informed consent. A leave-one-participant-out

Table 2 Operating parameters of the FMCW radar.

Parameter	Value
Starting frequency	75 GHz
Frequency modulation slope	64 MHz/ μ s
Sampling rate	4000 ksps
Number of chirps	16
Number of sampling points	128
Sampling time	32 μ s
Bandwidth	3.648 GHz
Frame period	50 ms

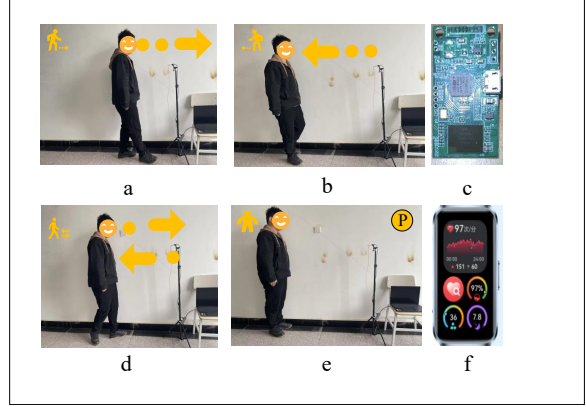


Fig. 3 Experimental platform and representative motion scenarios. (a) Forward motion; (b) backward motion; (c) radar acquisition board based on the STM32G473CCU6 microcontroller and RC7711B-AIP radio transceiver; (d) back-and-forth walking; (e) stationary standing; and (f) Huawei smart bracelet for synchronized reference heart-rate measurement.

protocol was adopted, with seven participants used for training and one for testing in each fold. During data collection, each participant wore a reference smart bracelet and performed standing, forward stepping, and backward stepping. For each participant, 1000 static samples and 300 dynamic range-time samples were collected, yielding 10,400 samples in total. All experiments were conducted in a closed corridor environment, as shown in Fig. 3.

4.2 Dual-Trajectory Experiment

Since no standardized quantitative metric is available for trajectory fitting in millimeter-wave vital-sign signals, we provide a qualitative comparison with the speed joint matrix method in [13]. As shown in Fig. 4, the original signal trajectory

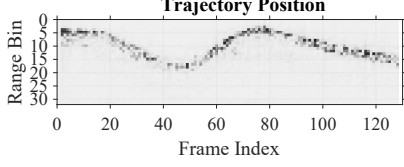


Fig. 4 Original target trajectory extracted from the range-time representation.

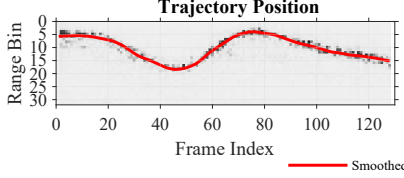


Fig. 5 Refined target trajectory after trajectory fitting.

contains missing segments, distortion, and ghosting. In contrast, the proposed method produces a clearer and more continuous trajectory, as shown in Fig. 5, and plausibly reconstructs missing segments. Further comparison in Fig. 6 shows that the speed joint matrix method is more sensitive to noise, resulting in distorted velocity estimates and inaccurate trajectory fitting, whereas the proposed method preserves trajectory continuity through multi-threshold local fitting and enhanced Kalman filtering[26].

The current validation is limited by the experimental scenarios and dataset scale. Further evaluation on larger datasets and in more diverse environments is needed to verify the generalizability of the proposed method.

4.3 Contrast Experiment

The proposed method was evaluated using mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), accuracy within ± 5 BPM, and Pearson correlation coefficient, all computed with respect to the wristband-based reference heart rate.

For conventional signal-processing baselines, we retain eight representative methods, namely band-pass filtering (BPF), FFT, wavelet transform (WT), EMD, VMD, Kalman filtering (KF), median filtering (MF), and singular spectrum analysis (SSA). These methods cover classical filtering, spectral estimation, time-frequency analysis, adaptive decomposition, state-space filtering, and robust smoothing. The estimated heart-rate values are reported in Table 3, while the summary

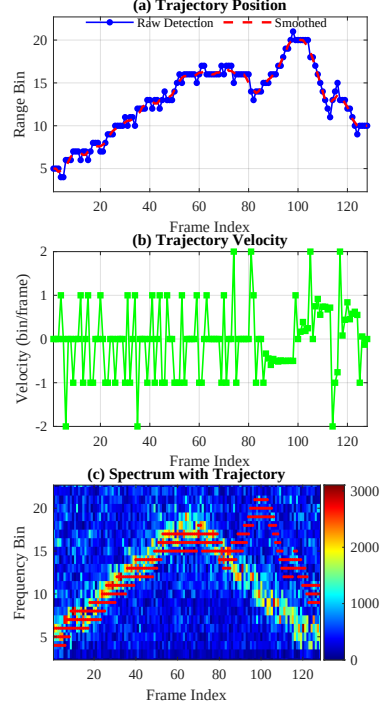


Fig. 6 Range-velocity-assisted trajectory refinement. (a) Range-domain fitting result. (b) Velocity-domain fitting result. (c) Comparison of the original and fitted bright-spot trajectories.

metrics are listed in Table 4. Most conventional methods are sensitive to motion-induced phase distortion and exhibit large estimation errors. Among them, MF performs relatively better, likely due to its stronger suppression of impulse-like artifacts. Nevertheless, the proposed framework still achieves the best overall agreement with the reference heart rate across all reported metrics.

We further compare the proposed method with representative deep learning models, including HRConvNet[18], HRNet, MobileNetV3-L[27], and ResNet[28]. The estimated heart-rate values are reported in Table 5, and the corresponding summary metrics are listed in Table 6. The proposed method consistently outperforms these baselines, indicating that its advantage comes not only from neural representation learning, but also from the combination of motion-aware phase extraction and structured waveform reconstruction.

Table 3 Estimated heart-rate values of representative conventional signal-processing baselines.

Trial	Estimated Heart Rate (bpm)									
	HR _{ref}	BPF	FFT	WT	EMD	VMD	KF	MF	SSA	Proposed
1	82	67.40	69.75	67.95	94.95	103.10	67.40	88.50	52.70	85.85
2	80	87.50	116.40	91.40	116.80	116.80	91.40	93.30	116.80	82.50
3	81	112.90	122.23	122.27	113.67	122.63	103.83	113.67	121.83	82.70
4	90	109.00	107.83	108.60	81.60	112.90	67.20	87.50	108.60	89.10
5	84	89.80	112.10	89.80	103.50	90.63	65.23	72.27	61.37	85.85
6	85	102.63	93.74	102.79	92.56	87.90	80.18	90.23	102.90	86.80
7	89	102.63	93.74	102.79	92.58	87.90	80.17	90.23	102.90	90.85

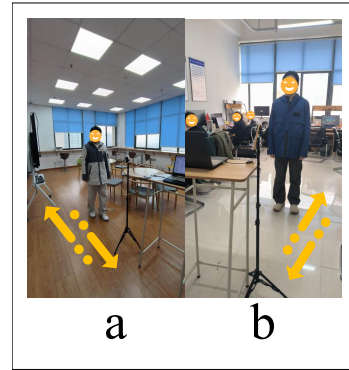
Table 4 Comparison with representative conventional signal-processing baselines.

Method	MAE	RMSE	MAPE (%)	A _{±5BPM} (%)	Pearson <i>r</i>
BPF	15.72	17.64	18.66	0.00	0.44
FFT	21.33	25.02	25.74	14.29	-0.17
WT	17.53	20.41	20.96	0.00	0.27
EMD	17.36	21.09	21.09	14.29	-0.85
VMD	19.01	24.23	23.02	28.57	-0.43
KF	14.86	16.20	17.64	14.29	-0.50
MF	10.45	14.44	12.74	28.57	-0.34
SSA	25.71	27.39	30.92	0.00	0.07
Proposed	2.06	2.23	2.47	100.00	0.94

4.4 Ablation Experiment

The proposed VCLA model is evaluated using mean squared error (MSE), mean absolute error (MAE), and signal-to-noise ratio (SNR). Each input segment contains 128 points, corresponding to 6.4 s and approximately 3–12 heartbeats. Before network inference, the signal is decomposed into five intrinsic mode functions using VMD, with the number of modes set to 5, the regularization parameter set to 2000, and the tolerance set to 1×10^{-7} .

To analyze the contribution of different architectural components, we compare single-branch, dual-branch, and full-fusion variants within a unified reconstruction setting. CNN only, LSTM only, and Autoformer only represent three single-branch baselines. CNN + LSTM, CNN + Autoformer, and LSTM + Autoformer evaluate pairwise combinations. CLA denotes the CNN–LSTM–Autoformer model with parallel fusion, while VCLA denotes the full model with VMD-enhanced structured input and adaptive fusion. We also include the pre-trained large model TimeXL as a representative foundation-model baseline. As summarized in Table 7, the complete VCLA framework achieves the lowest MSE

**Fig. 7** Representative evaluation scenarios used for cross-scenario validation: (a) single-person interference scenario and (b) multiple-person interference scenario.

and MAE as well as the highest SNR, confirming the effectiveness of combining structured decomposition, hierarchical temporal modeling, and adaptive fusion.

4.5 Robustness Evaluation

To evaluate cross-scenario generalization, the model was trained only on corridor data and tested in two unseen environments with different interference levels: a single-person meeting room

Table 5 Estimated heart-rate values of deep learning methods.

Trial	Estimated Heart Rate (bpm)					
	HR _{ref}	HRConvNet	HRNet	MobileNetV3-L	ResNet	Proposed
1	82	99.10	99.90	99.80	101.20	85.85
2	80	101.20	99.90	97.60	101.30	82.50
3	81	100.30	98.70	101.30	100.10	82.70
4	90	98.50	99.30	99.90	99.40	89.10
5	84	99.60	100.00	100.80	100.80	85.85
6	85	98.80	98.60	99.70	99.80	86.80
7	89	97.40	99.70	100.90	99.50	90.85

Table 6 Performance comparison of deep learning models.

Model	MAE	RMSE	MRE (%)	Acc _{±3BPM} (%)	R ²
HRConvNet	14.84	15.54	17.84	0.00	-17.85
HRNet	15.01	15.45	17.99	0.00	-17.63
MobileNetV3-L	15.57	15.93	18.63	0.00	-18.80
ResNet	15.87	16.42	19.04	0.00	-20.03
Proposed	2.06	2.23	2.47	85.70	0.61

Table 7 Ablation study on the contribution of individual components within the VCLA framework.

Model Architecture	MSE ↓	MAE ↓	SNR (dB) ↑
CNN only	0.043577	0.132870	8.07
LSTM only	0.032398	0.104496	9.36
Autoformer only	0.112487	0.222548	3.95
CNN + LSTM	0.021049	0.084587	11.23
CNN + Autoformer	0.036342	0.117032	8.86
LSTM + Autoformer	0.030486	0.099510	9.62
CLA (Parallel)	0.024639	0.091073	10.55
TimeXL	0.043500	0.128100	8.05
VCLA (Proposed)	0.019837	0.082578	11.49

Note: ↓ indicates a lower value is better; ↑ indicates a higher value is better. CLA: CNN–LSTM–Autoformer with parallel fusion. VCLA: Variational-mode-decomposition-enhanced CNN–LSTM–Autoformer with adaptive fusion. TimeXL denotes the pre-trained large-model baseline.

and a multi-person office. The test scenarios are shown in Fig. 7.

The proposed method achieves reliable heart rate estimation across both scenarios, with an overall MAE of 1.7 BPM. Specifically, the MAE is 0.5 BPM in the meeting room and 2.9 BPM in the office. The performance degradation in the multi-person setting is mainly due to the limitation of the data synthesis strategy, which models

self-motion interference but cannot fully reproduce complex multi-source disturbances caused by other moving individuals. Nevertheless, the method maintains sub-4 BPM accuracy under strong interference, demonstrating good robustness beyond the training conditions.

Compared with prior trajectory-fitting-based approaches that often rely on specialized hardware for obstacle discrimination, the proposed data-driven framework achieves robust performance with a more general hardware setup. However, the current validation remains limited by dataset scale, subject diversity, and the single-input-single-output radar configuration. Future work will focus on larger and more diverse cohorts, multi-person scenarios, and array-based sensing for improved contact-free monitoring.

5 Conclusion

This paper presents a robust framework for dynamic heart rate monitoring using FMCW radar. To address motion artifacts, a dual-trajectory fusion algorithm is first developed for adaptive range-gate tracking, enabling accurate estimation of time-varying chest displacement during motion. Based on this, a VCLA model is proposed to suppress motion-induced noise and recover the cardiogenic waveform.

Experimental results show that the proposed framework achieves reliable heart rate monitoring with a mean absolute error below 4 BPM, outperforming existing baseline methods in terms of accuracy and robustness. Cross-scenario evaluations further confirm its effectiveness in both single-person and multi-person environments, demonstrating its potential for contact-free vital sign monitoring in dynamic settings.

Declarations

Funding No external funding was received for this work.

Data availability statement The data and methods associated with this study will be continuously released and updated at the following GitHub repository: <https://github.com/jianantong38-spec/mmw>.

References

- [1] Lai, D., Cai, Z., Tian, G., *et al.*: Robust multi-person vital-sign sensing in indoor environments using FMCW MIMO radar. *Electronic Research Archive* **34**(3), 1477–1505 (2026)
- [2] Wu, Y., Ni, H., Mao, C., *et al.*: Non-intrusive human vital sign detection using mmWave sensing technologies: A review. *ACM Transactions on Sensor Networks* **20**(1), 1–36 (2023)
- [3] Cui, G., Wang, Y., Zhang, X., *et al.*: Non-contact heart rate variability monitoring with FMCW radar via a novel signal processing algorithm. *Sensors* **25**(17), 5607 (2025)
- [4] Qiao, X., Su, Y., Li, X., *et al.*: Millimeter-wave radar vital signs measurement with random body movement using missing data model. *IEEE Transactions on Instrumentation and Measurement* **74**, 4003814 (2025)
- [5] Mattos, A.B.R.C.D., Brante, G., Moritz, G.L., *et al.*: Classifying small animals, standing and lying down humans with multi mmWave radars. *IEEE Sensors Journal* **25**, 27513–27525 (2025)
- [6] Chang, H.-Y., Chang, R.Y., Chung, W.-H.: Bayesian-based weighted scheme for wireless vital sign estimation in small data conditions. *IEEE Sensors Journal* **25**(10), 17782–17794 (2025)
- [7] Sun, L., Huang, S., Li, Y., Gu, C., Pan, H., Hong, H., Zhu, X.: Remote measurement of human vital signs based on joint-range adaptive eemd. *IEEE Access* **8**, 68514–68524 (2020) <https://doi.org/10.1109/ACCESS.2020.2983694>
- [8] Singh, A., Rehman, S.U., Yongchareon, S., *et al.*: Human vital signs estimation using resonance sparse spectrum decomposition. *IEEE Transactions on Human-Machine Systems* **54**(3), 304–316 (2024)
- [9] Wang, W., Gong, Y., Zhang, H., *et al.*: Quantitative assessment of fall risk in the elderly through fusion of millimeter-wave radar imaging and trajectory features. *IEEE Access* **12**, 13370–13385 (2024)
- [10] Duan, X., Li, K., Liu, X., *et al.*: A non-contact sleep monitoring method based on millimeter-wave radar. *IEEE Transactions on Instrumentation and Measurement* **74**, 4006410 (2025)
- [11] Wang, S., Han, C., Guo, J., *et al.*: Mm-fgrm: Fine-grained respiratory monitoring using mimo millimeter wave radar. *IEEE Transactions on Instrumentation and Measurement* **73**, 1–13 (2023)
- [12] Marty, S., Ronco, A., Pantanella, F., *et al.*: Frequency matters: Comparative analysis of low-power FMCW radars for vital sign monitoring. *IEEE Transactions on Instrumentation and Measurement* **73**, 1–10 (2024)
- [13] Mercuri, M., Lorato, I.R., Liu, Y.H., *et al.*: Vital-sign monitoring and spatial tracking of multiple people using a contactless radar-based sensor. *Nature Electronics* **2**(6), 252–262 (2019)
- [14] Li, M., Yang, Y., Ke, X.: An adaptive heart rate monitoring system based on algebraic distance minimization. *IEEE Access*

- [15] Yang, G., Metcalfe, B.W., Watson, R.J., *et al.*: A noncontact cardiorespiratory monitoring system for drivers using millimeter wave radar with phase continuity tracking and signal quality index. *IEEE Internet of Things Journal* **12**(20), 43480–43494 (2025)
- [16] Wang, F., Zeng, X., Wu, C., *et al.*: mmHRV: Contactless heart rate variability monitoring using millimeter-wave radio. *IEEE Internet of Things Journal* **8**(22), 16623–16636 (2021)
- [17] Hao, Z., Gao, Y., Tang, Y., *et al.*: FMCW-based contactless heart rate monitoring. *Scientific Reports* **15**(1), 2616 (2025)
- [18] Zhao, W., Tong, G.: A deep learning based heart rate estimation method for millimeter wave radar. *Measurement* **255**, 117923 (2025)
- [19] He, Y., Wang, J., Li, Y., *et al.*: Dual-path CNN-BiLSTM for mmWave-based human skeletal pose estimation. *IEEE Sensors Journal* **25**, 11683–11696 (2025). Early access
- [20] Guo, Z., Zhang, B., Chen, J., *et al.*: Contactless fine-grained cardiac events detection and segmentation with radio frequency signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* **9**(3), 1–28 (2025)
- [21] Liu, K.C., Peng, S.Y., Tsao, Y., *et al.*: A cross-modal autoencoder for contactless electrocardiography monitoring using frequency-modulated continuous wave radar. *IEEE Sensors Journal* **24**(24), 41462–41473 (2024)
- [22] Wu, E.K., Fan, Q.G., Li, M.C., *et al.*: Non-contact monitoring of human cardiorespiratory activity during sleep using FMCW millimeter wave radar. *Measurement* **242**, 116144 (2025)
- [23] Jin, L., Pu, J., Li, H.: FMCW radar-based heartbeat recognition using SE-DenseNet for vital signs monitoring. *AEU-International Journal of Electronics and Communications* **200**, 155906 (2025) <https://doi.org/10.1016/j.aeue.2025.155906>
- [24] Bauder, C., Moadi, A.K., Rajagopal, V., *et al.*: Mm-mure: mmwave-based multi-subject respiration monitoring via end-to-end deep learning. *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology* **9**(1), 49–61 (2025)
- [25] Zhu, H., Zhao, Y., Zhang, Z., *et al.*: Multidimensional physiology-inspired enhanced vital sign monitoring using MIMO mmWave bio-radar. *IEEE Transactions on Instrumentation and Measurement* (2026)
- [26] Hao, Z., Wang, Y., Li, F., *et al.*: Detection of vital signs based on millimeter wave radar. *Scientific Reports* **15**(1), 28112 (2025)
- [27] Howard, A., Sandler, M., Chen, B., Wang, W., Chen, L.-C., Tan, M., Chu, G., Vasudevan, V., Zhu, Y., Pang, R., Adam, H., Le, Q.V.: Searching for MobileNetV3. In: *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 1314–1324 (2019). <https://doi.org/10.1109/ICCV.2019.00140>
- [28] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778 (2016). <https://doi.org/10.1109/CVPR.2016.90>