

1 Device-Free Respiration and Heartbeat Estimation with mmWave Radar

Foo Tuan Tong^a

^aGaitMetrics

This innovation introduces a contactless vital sign monitoring system utilizing Frequency-Modulated Continuous Wave (FMCW) radar technology. The method in this innovation can detect the heartbeat and respiration simultaneously using one deep neural network. It includes a hybrid architecture of Convolutional Neural Networks (CNNs) and Transformers for feature extraction and a novel customized diffusion model for estimation refinement. Through multi-task learning, the innovation utilizes the correlation between heartbeat and respiration signals and achieves the detection of these two biomarkers. The evaluation showcases precise detection of key ECG points while significantly reducing signal noise, promising groundbreaking advancements in non-invasive vital sign monitoring for health-care and daily life scenarios, with potential implications for remote patient care and health monitoring.

1 Patent Claims

Claim 1: A contactless vital sign monitoring system utilizing Frequency-Modulated Continuous Wave (FMCW) radar technology, comprising:

1. A radar sensor configured to emit electromagnetic waves in a continuous wave manner and receive reflected waves from a subject;
2. A deep neural network comprising: (a) A hybrid architecture of Convolutional Neural Networks (CNNs) and Transformers for feature extraction; and (b) A customized diffusion model for estimation refinement;

A report for patent application

3. Said deep neural network further configured to: (a) Simultaneously detect heartbeat and respiration signals from the reflected waves; (b) Utilize multi-task learning to exploit the correlation between the detected heartbeat and respiration signals;
4. The system further configured to: (a) Precisely detect key points on an Electrocardiogram (ECG); and (b) Significantly reduce signal noise during the detection process;
5. The system providing groundbreaking advancements in non-invasive vital sign monitoring for healthcare and daily life scenarios;
6. The system having potential implications for remote patient care and health monitoring.

2 Invention Background

The monitoring of vital signs plays a pivotal role in the field of healthcare (Mok et al. 2015). Real-time and continuous monitoring of crucial vital signs, such as heart rate and respiration, holds the potential to enable early disease detection and serves as a vital reference for healthcare professionals in diagnosing specific conditions. Anomalies in heart rate or breathing patterns frequently serve as early indicators of various significant illnesses. Furthermore, the stability of an individual's physical condition is often reflected in the normalcy of their vital signs (Cardona-Morrell et al. 2016), making accurate and dependable vital sign monitoring indispensable in the daily lives of ordinary individuals. This is especially true for those with unstable physical conditions, including the elderly, recovering patients, or individuals with chronic illnesses.

Considerable progress has been made in the realm of vital sign monitoring. In the medical domain, numerous methods are available for monitoring vital signs. The conventional approach to measuring the electrocardiogram (ECG) involves utilizing an electrocardiogram machine (Serhani et al. 2020). In cases of severe cardiac conditions or the need for prolonged and continuous cardiac monitoring, more invasive methods such as Implantable Cardioverter Defibrillators (ICDs) are employed. Whereas, respiration is a vital sign that is more readily observable. In clinical settings, the most prevalent methods for monitoring respiration encompass visual assessment by medical practitioners or the use of a stethoscope (Proulx 1999). For advanced medical monitoring, common indicators include pulse oximetry (SpO₂) and end-tidal carbon dioxide (EtCO₂) levels. Despite their proven reliability, these methods possess inherent limitations.

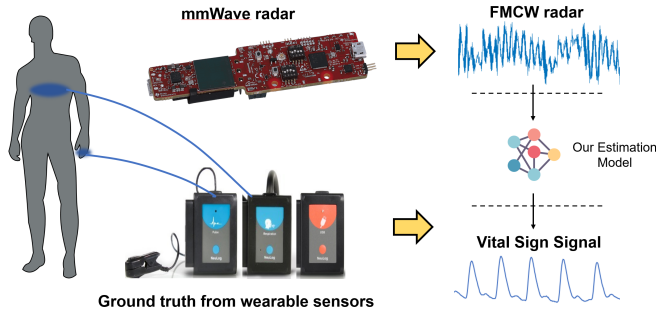


Figure 1: Contactless vital sign detection System Overview

The primary impediment to their widespread adoption in daily life lies in their heavy reliance on specialized medical facilities, professional-grade equipment, and trained healthcare personnel (Schläpfer & Wellens 2017). These medical devices are expensive and typically necessitate the expertise of professionals or even surgical procedures.

3 Challenges in Non-contact Vital Sign Monitoring

In recent years, the use of RF sensors has emerged as one of the most promising non-contact methods for monitoring vital signs. Many studies have successfully employed RF sensors to capture respiratory rate and heart rate through sophisticated signal analysis techniques (Vinci et al. 2013, Tu & Lin 2015). However, raw radar data poses challenges due to its high dimensionality and low signal-to-noise ratio (SNR), making the application of traditional signal analysis methods to process the data and extract precise, undistorted information a formidable task. The noise in radar data emanates from various sources, i.e., multipath effects and unrelated movements within the monitored space apart from chest movement. Furthermore, Even the heartbeat and respiration, both crucial vital signs, can introduce interference with each other’s detection tasks. Radar signals extract vital sign information by detecting the up-and-down movement of the human chest. The long sequence in vital sign monitoring is also a challenge as deep models are normally not good at predicting long sequence.

To address these challenges and predict fine-grained ECG signals using FMCW radar data, we propose to eliminate the noise in radar signals by a novel CNN-transformer prediction network and a diffusion model that further refines the vital sign signals. We enable the network to learn both respiration and heart

beat simultaneously. Consequently, respiratory and cardiac information are intertwined and reflected within the coherent radar data. The proposed system can estimate vital sign information from the raw radar signals very accurately.

The novelties of this innovation can be summarized as follows:

- **Design of Contactless Vital Sign Monitoring Device:** In contrast to traditional contact-based methods, our device excels in long-term monitoring during daily life. Its reduced invasiveness ensures a more comfortable experience for users, particularly those with special medical needs.
- **Transformer-Diffusion Net with Multi-Task Learning:** Based on a comprehensive analysis of the radar signals, we propose a Transformer and diffusion model capable of concurrently predicting respiratory and ECG vital sign parameters. Our innovation successfully extracts precise vital sign from noisy radar data.
- **Experimental Validation of Vital Sign Extraction from FMCW Radar:** Through a series of experiments, we substantiate the efficacy of our method in extracting vital signs from FMCW radar signals.

4 Inention Methodology

4.1 Reflected Signals from FMCW Radar

Before introducing our methodology, it is essential to provide a concise explanation of how radar data can be used to reconstruct vital signs (Li et al. 2019). FMCW radar emits a sawtooth wave signal, represented by Equation 1.1:

$$S_T(t) = \cos \left[2\pi \left(f_0 t + \frac{\mu t^2}{2} + \phi_0 \right) \right]. \quad (1.1)$$

Equation 1.2 illustrates the echo signal expression reflected by an object at a distance R with velocity V (positive in the direction close to the radar):

$$S_R(t) = K_r A_0 \cos \left[2\pi \left(f_0(t - \tau) + \frac{\mu(t - \tau)^2}{2} + \phi_0 + \vartheta_0 \right) \right], \quad (1.2)$$

Here, f_0 represents the center frequency of the transmitted signal (i.e., the instantaneous frequency of the transmitted signal at $t = 0$), μ is the frequency modulation slope, ϕ_0 is the initial phase when the signal is transmitted, and ϑ_0 accounts for additional phase introduced by target reflection. $\tau = \frac{2(R-vt)}{c}$ denotes

the echo delay of the signal. By mixing the received echo signal with the transmitted signal, we obtain a beat signal presented in Equation 1.3:

$$S(t) = A \cos \left[2\pi \left(f_0 \frac{2R}{c} + \left(\frac{2\mu R}{c} - 2 \frac{v f_0}{c} \right) t \right) + \phi_0 \right]. \quad (1.3)$$

Consequently, the reflected signal encapsulates micro-displacement and velocity information resulting from heartbeat and respiration. This implies that the radar's reflected signal from the chest cavity inherently contains vital sign information.

For FMCW radar, its range resolution is defined as per Equation 1.4:

$$R_{res} = \frac{c}{2f_c}, \quad (1.4)$$

Where f_c denotes the frequency modulation bandwidth of the radar. The larger the frequency modulation bandwidth, the more sensitive the carrier phase becomes, making it more responsive to changes in the reflected signal due to displacement.

It's worth noting that ECG reflects potential differences resulting from polarization and depolarization, while FMCW radar captures mechanical movements. Although these two signal types reflect distinct physical phenomena, both describe the same event. Previous research has established the existence of a modelable mapping relationship between cardiac mechanical movements and cardiac electrical activity (Bers 2002, Gurev et al. 2011). Hence, in theory, predicting ECG using radar data is feasible.

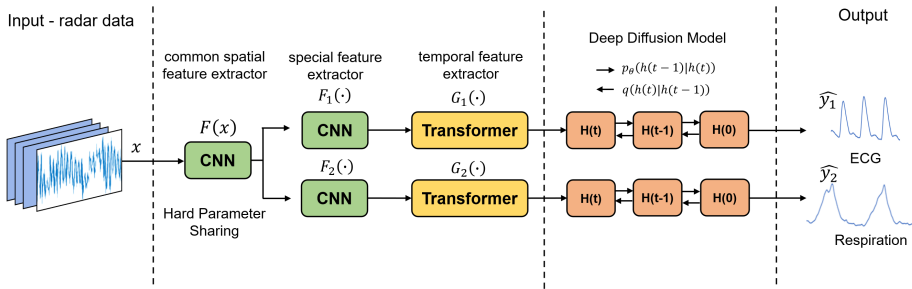


Figure 2: MultiVSNet architecture for vital signs prediction.

4.2 Estimating Vital Signs from Radar Signals

Our approach to vital sign prediction involves employing deep learning networks to establish the mapping relationship between mmwave signals and vital signs. However, traditional convolutional networks (CNN) focus locally, limiting their ability to grasp global correlations in long sequences, while recurrent neural networks (RNN) can suffer from vanishing or exploding gradients, necessitating increased computational and storage resources. To address these challenges, we propose to leverage CNN and Transformer layers for feature extraction first. The radar data undergoes initial processing in CNN blocks, extracting spatial features. Then follows the Transformer encoding layers for further processing. After decoding, we further input the estimation into a diffusion model for further refinement. The decoder employs two linear layers, and positional relationships within the sequence are represented using sinusoidal positional encoding. The whole network structure is summarized in Table 1.

Table 1: Multi-Task Learning Network Parameters.

	Layer Name	AET (ms)
2*Bottom Layer	$conv1d \times 3$	$8 \times 8, stride = 2$
	$max - pooling \times 3$	$4 \times 4, stride = 2$
5*Top Layer	$conv1d \times$	$8 \times 8, stride = 0$
	$max - pooling \times 3$	$4 \times 4, stride = 2$
	3*transformer	$d_{model} = 512, d_{ff} = 256$
		$d_k = d_q = d_v = 64$
		$heads = 8, layer = 2$
Decoder	fully connected layer	number of neurons = 256

4.2.1 Learning Framework

To enhance the robustness of our system and make effective use of the data, we employ a multitask learning approach. As shown in Fig. 2, we simultaneously train the network to predict human respiration and ECG. These auxiliary tasks serve as regularization terms, facilitating the learning of more standardized features (Ruder 2017). This enables the network to focus on meaningful information and disregard noise, even in data with low signal-to-noise ratio (SNR). Our multitask learning framework employs shared parameters in the feature extractor, which includes CNN and max-pooling layers. This shared feature extractor allows the network to learn a common representation for both primary and auxiliary tasks. It encourages the network to share early representations with the

same parameters. For the top-level generation task, a CNN block and a Transformer block are utilized for sequence prediction. Each task is addressed by a separate branch, with distinct parameters for the top-level networks, enabling specialization for each task.

Spatial-Temporal Feature Extraction. While predicting human respiration and ECG may appear as distinct tasks, they share a common spatial component. Both vital signs are reflected in radar signals from the chest’s spatial position, collectively representing an individual’s health. This interrelationship implies some similarities between the spatial patterns associated with heartbeats and respiration. To capture these shared patterns, we employ a shared CNN module for spatial feature extraction. Then we utilize the Transformer network that proves effective in handling long sequence prediction tasks. Its self-attention mechanism allows for the simultaneous consideration of information from various positions in the sequence. Thus, the Transformer captures both local and global relationships within the sequence. Considering the distinct waveform periods and patterns of ECG and respiration, we employ two independent and parallel Transformer networks to extract temporal features for each subtask. These temporal features serve as inputs to the decoder.

Diffusion Models for Vital Sign Refinement.

Diffusion models are a class of probabilistic models used in deep learning to describe the temporal evolution of data. In your context, you have input data denoted as $H(0)$, representing the initial state of the heartbeat signal, and $H(t)$ representing the data at timestep t , which is subject to noise. A diffusion model aims to refine this signal and obtain a cleaner output.

In practice, we often work with discrete-time models for simplicity. Discretizing the above continuous-time process, we get:

$$H(t + \Delta t) = H(t) + \Delta t \cdot f(H(t), t; \theta) + \sqrt{\Delta t \cdot \beta} \cdot \epsilon(t), \quad (1.5)$$

where Δt is the time step.

Diffusion models are trained using variational inference to estimate the model parameters θ . The goal is to learn a distribution over the data at each time step conditioned on the initial data $H(0)$:

$$p(H(t)|H(0), \theta) = \int p(H(t)|H(0), \theta)q(H(0))dH(0), \quad (1.6)$$

where $q(H(0))$ is an approximation to the true posterior distribution over the initial data. By iteratively refining the data using the diffusion process and optimizing θ using variational inference, the model can produce samples that match

the observed data after a finite number of time steps, effectively denoising the vital sign signal.

4.3 Objectives and Optimization

4.3.1 Training Feature Extractor

We first train the prediction network with the CNN and transformer. In this formulation, the loss function is expressed as Equation 1.7:

$$\mathcal{L}_{total} = \mathcal{L}_1(\omega_1) + \mathcal{L}_2(\omega_2), \quad (1.7)$$

where $\mathcal{L}_1$ represents the primary task loss, $\mathcal{L}_2$ denotes the auxiliary task loss, and $\mathcal{L}_{total}$ represents the final loss function value. The parameters ω_1 , and ω_2 are the network parameters to be optimized.

4.3.2 Training Diffusion Model

Then we train the diffusion model. The training procedure for a diffusion model involves learning the model parameters θ using variational inference. Here's a high-level overview of the training steps:

1. Initialize the model parameters θ .
2. Given a batch of initial data samples $\{H(0)_i\}_{i=1}^N$, simulate the diffusion process to generate noisy data samples $\{H(t)_i\}_{i=1}^N$ at multiple time steps t .
3. Compute the negative log-likelihood loss between the noisy data and the true data at each time step.
4. Compute the variational inference loss, typically based on the evidence lower bound (ELBO), to approximate the posterior distribution $p(H(0)|H(t), \theta)$.
5. Update the model parameters θ using gradient-based optimization methods, such as stochastic gradient descent (SGD), to minimize the combined loss.
6. Repeat the above steps for multiple epochs until convergence.

The loss function used during training typically consists of two components: the negative log-likelihood loss and the variational inference loss. The negative

log-likelihood loss at time step t measures the dissimilarity between the generated data $H(t)$ and the true data. It is often defined as the Gaussian log-likelihood loss:

$$\mathcal{L}_{\text{NLL}}(H(t), \hat{H}(t)) = \frac{1}{2\sigma^2} \|H(t) - \hat{H}(t)\|^2, \quad (1.8)$$

where:

- $H(t)$ is the true data at time step t .
- $\hat{H}(t)$ is the generated data at time step t using the diffusion process.
- σ is a hyperparameter controlling the scale of the Gaussian distribution.

The variational inference loss, often based on the ELBO, is used to approximate the posterior distribution $p(H(0)|H(t), \theta)$. It encourages the model to learn meaningful representations of the initial data. The ELBO loss can be defined as:

$$\mathcal{L}_{\text{VI}} = \mathbb{E}_{q(H(0))} \left[\log \frac{p(H(0), H(t)|\theta)}{q(H(0))} \right], \quad (1.9)$$

where:

- $q(H(0))$ is the approximate posterior distribution over the initial data.
- $p(H(0), H(t)|\theta)$ is the joint distribution over the initial data and the data at time step t given the model parameters θ .

The total loss used during training is a combination of the negative log-likelihood loss and the variational inference loss, typically with a hyperparameter weighting factor:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{NLL}} + \lambda \cdot \mathcal{L}_{\text{VI}}, \quad (1.10)$$

where λ controls the trade-off between matching the observed data and approximating the posterior distribution.

5 Experiments

5.1 Setup

We utilized the IWR6843AOPEVM board for ECG detection (Texas Instruments n.d.[b]). Subjects remained seated with the radar placed approximately 0.5 meters

from their chests. The IWR6843AOPEVM was combined with the DAC1000 (Texas Instruments n.d.[a]) to gather raw ADC data from the radar. This radar operates within the 60 to 64 GHz frequency range, featuring four receivers and three transmitters.

Configuration files were sent to the radar via Python and TI's mmwave studio to configure the DAC1000 for ADC data collection. Ground truth data for heartbeat and respiration were concurrently collected using the NeuLog heartbeat (NeuLog n.d.[a]) and respiration waveform measurement device (NeuLog n.d.[b],[c]). This portable electrocardiogram generator employed finger electrodes for ECG measurement. Timestamps were added during data collection for subsequent data alignment, creating a suitable dataset for supervised learning.

The dataset comprises data from four healthy subjects (two females and two males) in an environment with furniture that introduced multipath effects and increased radar data noise. It includes 48,000 frames of valid radar data collected at a rate of 25 frames per second. The corresponding respiration and ECG data contained around 3,000 heartbeat samples. Samples were segmented into 64-frame windows, with 70% of the data used for training and the remaining 30% reserved for testing.

5.2 Evaluation Metrics

In the context of ECG analysis for diagnosing cardiac electrical activity and myocardial function, the focus lies on morphological waveform features, specifically the characteristics of individual peaks and inter-peak intervals. These features are vital for determining the rhythm of cardiac electrical activity. The signal's amplitude is of lesser importance. Hence, the network's performance is evaluated based on the accuracy of detected time points for R-peaks and T-peaks. The R-peak is defined as the maximum value identified within a complete ECG period, while the T-peak is the local maximum value appearing after the R-peak. The detection of these critical points serves as the basis for evaluation. To quantitatively assess the reconstructed waveform's performance, the average absolute error between the time points of these peaks in the reconstructed and actual ECG signals is calculated. This error is represented as the R-peak prediction error (AER) and the T-peak prediction error (AET). For each signal cycle, AER and AET are calculated as follows:

$$AER = \frac{1}{Nr} \sum_{n=1}^N |RP_{true}(n) - RP_{pred}(n)|, \quad (1.11)$$

$$AET = \frac{1}{Nr} \sum_{n=1}^N |TP_{true}(n) - TP_{pred}(n)|, \quad (1.12)$$

Here, N represents the total number of sampling points in the evaluated signal segment, and n denotes the n th cycle detected within the sequence under evaluation. $RP_{true}(n)$ and $TP_{true}(n)$ correspond to the sequence positions of detected R and T peaks in the actual ECG signal, while $RP_{pred}(n)$ and $TP_{pred}(n)$ denote the sequence positions of R and T peaks detected in the reconstructed ECG signal. The parameter r represents the signal sampling rate, converting differences in sequence positions into time errors.

5.3 Experimental Results

The results are shown in the Table 2. It can be observed that our method achieves smaller AER and AET values, thus better capturing the occurrence times of the R peak and T peak. The predicted waveforms generally exhibit a good representation of the actual fluctuations, as shown in Fig. 3.

Table 2: The AERs, AETs of baseline and our method.

Network	AER (ms)	AET (ms)
CNN and LSTM hybrid network(Yamamoto et al. 2020)	15.7	23.8
CNN model(Ha et al. 2020)	28.8	92.8
CNN and GRU hybrid network	12.6	17.4
LSTM model	29.0	25.8
Our Method	8.2	9.8

References

- Bers, Donald M. 2002. Cardiac excitation–contraction coupling. *Nature* 415(6868). 198–205.
- Cardona-Morrell, M, M Prgomet, RM Turner, M Nicholson & K Hillman. 2016. Effectiveness of continuous or intermittent vital signs monitoring in preventing adverse events on general wards: a systematic review and meta-analysis. *International journal of clinical practice* 70(10). 806–824.

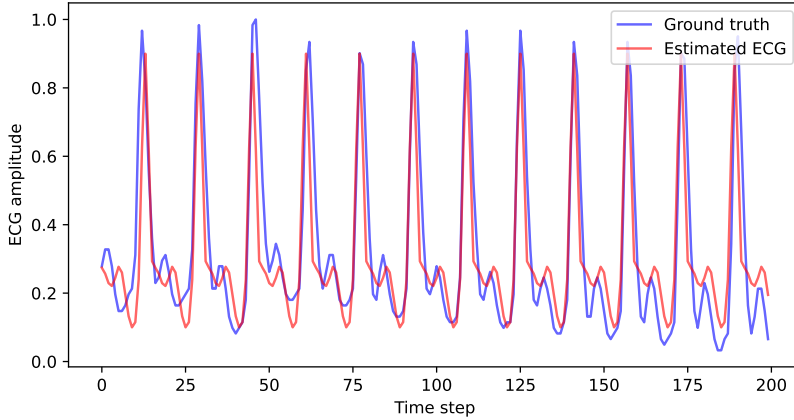


Figure 3: The prediction results of the innovation.

- Gurev, Viatcheslav, Ted Lee, Jason Constantino, Hermenegild Arevalo & Natalia A Trayanova. 2011. Models of cardiac electromechanics based on individual hearts imaging data: image-based electromechanical models of the heart. *Biomechanics and modeling in mechanobiology* 10. 295–306.
- Ha, Unsoo, Salah Assana & Fadel Adib. 2020. Contactless seismocardiography via deep learning radars. In *Proceedings of the 26th annual international conference on mobile computing and networking*, 1–14.
- Li, Xinyu, Yuan He & Xiaojun Jing. 2019. A survey of deep learning-based human activity recognition in radar. *Remote Sensing* 11(9). 1068.
- Mok, Wen Qi, Wenru Wang & Sok Ying Liaw. 2015. Vital signs monitoring to detect patient deterioration: an integrative literature review. *International journal of nursing practice* 21. 91–98.
- NeuLog. N.d.(a). *Heart Rate & Pulse logger sensor NUL-208*. 2017. <https://neulog.com/heart-rate-pulse/>.
- NeuLog. N.d.(b). *Respiration Monitor Belt logger sensor NUL-236*. 2017. <https://neulog.com/respiration-monitor-belt/>.
- NeuLog. N.d.(c). *USB module USB-200*. 2017. <https://neulog.com/usb/>.
- Proulx, Jeffrey. 1999. Respiratory monitoring: arterial blood gas analysis, pulse oximetry, and end-tidal carbon dioxide analysis. *Clinical Techniques in Small Animal Practice* 14(4). 227–230.
- Ruder, Sebastian. 2017. An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*.

- Schläpfer, Jürg & Hein J Wellens. 2017. Computer-interpreted electrocardiograms: benefits and limitations. *Journal of the American College of Cardiology* 70(9). 1183–1192.
- Serhani, Mohamed Adel, Hadeel T. El Kassabi, Heba Ismail & Alramzana Nujum Navaz. 2020. Ecg monitoring systems: review, architecture, processes, and key challenges. *Sensors* 20(6). 1796.
- Texas Instruments. N.d.(a). *DCA1000EVM*. 2023. <https://www.ti.com/tool/DCA1000EVM?keyMatch=DAC1000>.
- Texas Instruments. N.d.(b). *IWR6843AOPEVM*. 2023. <https://www.ti.com/product/IWR6843AOPEVM/part-details/IWR6843AOPEVM>.
- Tu, Jianxuan & Jenshan Lin. 2015. Fast acquisition of heart rate in noncontact vital sign radar measurement using time-window-variation technique. *IEEE Transactions on Instrumentation and Measurement* 65(1). 112–122.
- Vinci, Gabor, Stefan Lindner, Francesco Barbon, Sebastian Mann, Maximilian Hofmann, Alexander Duda, Robert Weigel & Alexander Koelpin. 2013. Six-port radar sensor for remote respiration rate and heartbeat vital-sign monitoring. *IEEE Transactions on Microwave Theory and Techniques* 61(5). 2093–2100.
- Yamamoto, Kohei, Ryosuke Hiromatsu & Tomoaki Ohtsuki. 2020. Ecg signal reconstruction via doppler sensor by hybrid deep learning model with cnn and lstm. *Ieee access* 8. 130551–130560.