DeepVS: A Deep Learning Approach For RF-based Vital Signs Sensing

Zongxing Xie
Stony Brook University
zongxing.xie@stonybrook.edu

Hanrui Wang
Massachusetts Institute of Technology
hanrui@mit.edu

Song Han
Massachusetts Institute of Technology
songhan@mit.edu

Elinor Schoenfeld
Stony Brook University
elinor.schoenfeld@stonybrook.edu

Fan Ye
Stony Brook University
fan.ye@stonybrook.edu

ABSTRACT

Vital signs (e.g., heart and respiratory rate) are indicative for health status assessment. Efforts have been made to extract vital signs using radio frequency (RF) techniques (e.g., Wi-Fi, FMCW, UWB), which offer a non-touch solution for continuous and ubiquitous monitoring without users' cooperative efforts. While RF-based vital signs monitoring is user-friendly, its robustness faces two challenges. On the one hand, the RF signal is modulated by the periodic chest wall displacement due to heartbeat and breathing in a nonlinear manner. It is inherently hard to identify the fundamental heart and respiratory rates (HR and RR) in the presence of higher order harmonics of them and intermodulation between HR and RR, especially when they have overlapping frequency bands. On the other hand, the inadvertent body movements may disturb and distort the RF signal, overwhelming the vital signals, thus inhibiting the parameter estimation of the physiological movement (i.e., heartbeat and breathing). In this paper, we propose DeepVS, a deep learning approach that addresses the aforementioned challenges from the non-linearity and inadvertent movements for robust RF-based vital signs sensing in a unified manner. DeepVS combines 1D CNN and attention models to exploit local features and temporal correlations. Moreover, it leverages a two-stream scheme to integrate features from both time and frequency domains. Additionally, DeepVS unifies the estimation of HR and RR with a multi-head structure, which only adds limited extra overhead (<1%) to the existing model, compared to doubling the overhead using two separate models for HR and RR respectively. Our experiments demonstrate that DeepVS achieves 80-percentile HR/RR errors of 7.4/4.9 beat/breaths per minute (bpm) on a challenging dataset, as compared to 11.8/7.3 bpm of a non-learning solution. Besides, an ablation study has been conducted to quantify the effectiveness of DeepVS.

CCS CONCEPTS

• **Applied computing → Bioinformatics**; **Health informatics**;
• **Computing methodologies → Neural networks**.

KEYWORDS

Vital Signs; RF; Deep Learning; CNN; Attention Mechanism

This work is supported in part by NSF grants 1951880, 2028952, 2119299.

BCB '22, August 7–10, 2022, Northbrook, IL, USA

ACM ISBN 978-1-4503-9386-7/22/08…$15.00
https://doi.org/10.1145/3535508.3545554

ACM Reference Format:
Zongxing Xie, Hanrui Wang, Song Han, Elinor Schoenfeld, and Fan Ye. 2022. DeepVS: A Deep Learning Approach For RF-based Vital Signs Sensing. In *13th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics (BCB '22), August 7–10, 2022, Northbrook, IL, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3535508.3545554

1 INTRODUCTION

Heart and respiration rates (HR, RR) are two fundamental vital signs for human health status assessment. While such vital signs are usually measured at hospital visits, people start to value the continuous vital signs data collected in individuals' home environment, that enable insightful profiling, detecting potential health problems, or tracking the progression of chronic diseases. Wearables (e.g., Apple Watch, Fitbit) are popular for continuous vital signs monitoring. However, they require frequent charging and wearing, adding to physical and cognitive challenges to compliance, especially among older adults, the population that may benefit the most from the continuous vital signs monitoring. To achieve comfortable solutions for longitudinal monitoring, researchers have explored RF-based vital signs sensing designs [8] including using Wi-Fi [13, 17], FMCW [2, 18], and UWB [15, 19]. Such work has demonstrated the feasibility of RF-based vital signs sensing. In addition, they do not expose imagery privacy nor require users to wear any device or pay cooperative efforts, promising for continuous in-home monitoring.

However, two key challenges remain in achieving robustness of RF-based vital signs sensing for broader applications in practical scenarios. First, because the received RF signal is composed of reflections from objects or body parts at different distances, the heartbeat and respiration signals are mixed in a nonlinear manner, thus hard to separate and identify their fundamental components due to in the presence of higher order harmonics and intermodulation between HR and RR [21]. Second, while RF signal is sensitive to the chest wall displacement for vital signs sensing, the large body movement will overwhelm the signal modulated by minute physiological movements, thus distorting the parameter estimation of HR/RR. This may happen frequently but unconsciously in realistic scenarios during vital signs monitoring. In practice, it is infeasible to find a an analytical model that directly characterizes the vital signs from the noisy RF signal given the aforementioned challenges from the non-linearity and the inadvertent body movement.

In this paper, we propose DeepVS, a hybrid deep learning scheme that combines 1D CNN [6] and attention models [16] to exploit local features and temporal correlations to address the aforementioned challenges from non-linearity and inadvertent movements for robust RF-based vital signs sensing in a unified manner. Moreover, it leverages a two-stream scheme to integrate features from both time and frequency domains. Additionally, DeepVS unifies the estimation of HR and RR with a multi-head structure, which only

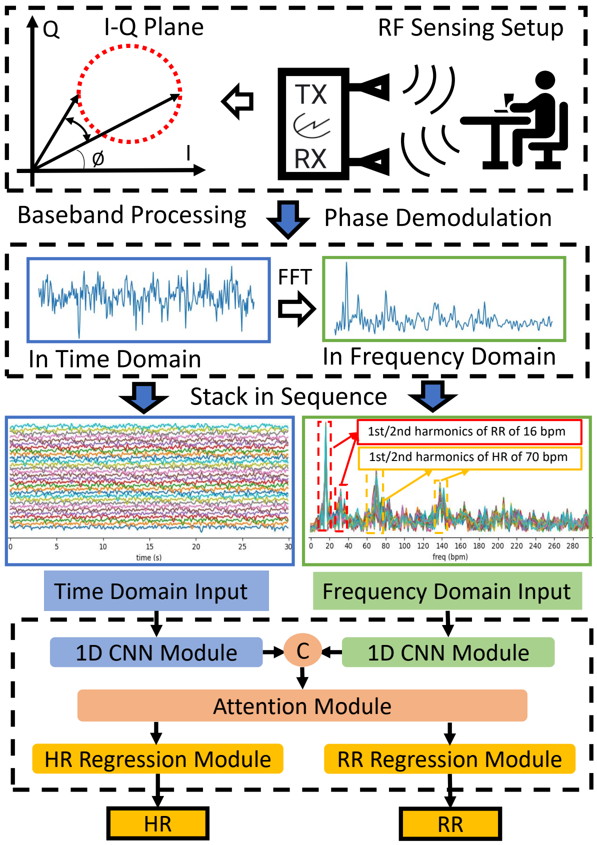


Figure 1: Overview of DeepVS framework. In a typical setup, the RF sensor emits RF signals into the environment and receive the echoes reflected by the subject in the field of view (FoV). By downconversion from RF, the baseband signal is obtained in the In-phase and quadrature (I-Q) plane for phase demodulation. The phase signal in the time domain changes periodically corresponding to the physiological movements, and several prominent frequency components can be easily detected in the frequency domain with Fourier Transform. Because the signal in a single time window may be disturbed and distorted from time to time, we stack signals of several consecutive time windows in a sequence as the input to DeepVS for vital signs estimation. DeepVS exploits 1D CNN to extract local features in both time and frequency domain in a two-stream scheme, the attention module to capture temporal correlations, and two regression layers in a multi-head scheme to estimate HR and RR respectively.

adds limited extra overhead (<0.1%) to the existing model, compared to doubling the overhead using two separate models for HR and RR respectively. Our experiments demonstrate that DeepVS reduces 80-percentile RR/HR errors by 37.3/32.8%, as compared to 11.8/7.3 bpm of a non-learning solution [20]. Besides, an ablation study has been conducted to quantify the effectiveness of DeepVS.

Specifically, we summarize our key contributions as follows:

- We propose a unified deep learning framework for robust RF-based vital signs sensing by combining CNN and attention models, integrating the local features and temporal correlations from time/frequency domains to address the challenges from non-linearity and body movements simultaneously.
- We evaluate DeepVS using 80k data samples collected over 1 month in typical home environments. Results show that DeepVS achieves 80-percentile RR/HR errors of 7.4/4.9 bpm on a challenging dataset with half the overhead compared to using two individual models for HR and RR separately.

2 BACKGROUND AND RELATED WORK OF RF-BASED VITAL SIGNS SENSING

In this section, we first describe the background of RF-based vital signs sensing with signal modeling and sources of challenges, and then discuss the related work.

2.1 Background

The rationale of RF-based vital signs sensing lies in that the minute chest wall displacement due to physiological movements (i.e., heartbeat and respiration) changes the propagation distance of the reflected RF signal, thus the phase of the received RF signal, from which we can extract vital signs. The one-way signal propagation distance can be modeled as:

$$\begin{aligned} d(t) &= d_0 + D(t) \\ &= d_0 + d_b(t) + d_r(t) + d_h(t), \end{aligned} \tag{1}$$

where d_0 is the initial distance between the RF sensor and the targeted chest wall; $D(t)$ is the whole displacement attributed to d_b, d_r and d_h, the variations by body movements, respiration and heartbeat respectively.

Ideally, when the body movement is negligible, the phase of the received RF signal is linearly modulated by the chest wall displacement, and can be modeled as:

$$\phi_D(t) = 2\pi f_c D(t)/c, \tag{2}$$

where f_c is the center frequency of the RF signal, and c is the speed of light. By extracting the frequency components from the modulated phase, we are able to extract heart and respiratory rates.

However, robust vital signs sensing is greatly challenged by the non-linearity and disturbance from inadvertent movements in realistic scenarios. Under non-linear channels, the perceived phase may be approximated by the Taylor series in terms of the displacement $D(t)$ as follows:

$$\phi_D(t) = \frac{2\pi f_c}{c} \sum_{i=1}^{\infty} a_i D^i = \frac{2\pi f_c}{c}(a_1 D(t) + a_2 D^2(t) + \ldots), \tag{3}$$

where a_i is the coefficient of the i-th order term. Higher order terms lead to the intermodulation effect, producing new components at frequencies that are the linear combination of HR and RR (i.e., $\{m\text{HR}\pm\ n\text{RR} | m, n \in \mathbb{N}_0\}$) and hard to be distinguished from true vital signs. The presence of the large body movements $d_b(t)$ would further complicate the problem of vital signs sensing, as it will be dominating $D(t)$ compared to the minute physiological movements.

2.2 Related Work

Since the concept of RF-based vital signs sensing was introduced back in the 1970's [12], it has been widely studied as new RF techniques emerge including UWB [15], FMCW [2], and Wi-Fi [13, 17]. However, the aforementioned challenges (in §2.1) from the non-linearity and inadvertent body movements remain serious obstacles for robustness in practical scenarios.

For the non-linearity challenge, not treated sufficiently in the literature, some effort from the electrical engineering community has demonstrated that certain spectral patterns can be leveraged to distinguish the vital signs components from the interference of intermodulation and noise. Heuristic-based methods [15] have been proposed to detect the "path", defined as a set of consecutive, approximately equally spaced spectral peaks, indicating the existence of harmonics, such that the frequency interval of the most powerful "path" is the estimate of vital signs. To further push the limit, a probabilistic weighting framework [20] was proposed to dynamically combine an ensemble of vital signs estimators, based on a set of heuristics in terms of the prominence of spectral peaks, the existence of harmonics, and the temporal locality of vital signs. However, heuristic methods may not always be able to capture the varying patterns.

For the challenge from the disturbance due to the body movements, proposals can be categorized into two types. One is detecting and excluding the distorted signal. Existing methods usually formulate it as a classification task [21], and devise different features such as spectral energy [2] or temporal waveform [8, 13] to assess and determine the signal quality thus availability for vital signs extraction. The other type is reconstructing the vital signal in the presence of body movements. While traditional blind source separation methods [4] usually fail to meet the end, early proposals with delicate placements of two radars [10, 11] showed encouraging results in handling the interference from body movements. However, realistic application scenarios in daily environments pose difficulties for such tricky placements to achieve ideal performance. Researchers reported promising results of vital signs reconstruction with recent advances in deep learning [3, 7, 22]. BreathListener [22] exploited a GAN architecture [5] to reconstruct the respiration waveform in driving environments. RF-SCG [7] introduced a 1D CNN model to reconstruct the heartbeat signal by optimizing the loss function of template matching. MoVi-Fi [3] leveraged a Contrastive Learning scheme [9] to achieve the nonlinear ICA [9] thus separating the signal of vital signs from the body movements. However, they usually focused on reconstructing either the heartbeat or respiration waveform but not the end-to-end estimation of both HR and RR in a unified manner.

3 DEEPVS FRAMEWORK

In this section, We introduce DeepVS, a deep learning framework for reliable vital signs estimation facing the challenges of nonlinearity and body movements in a unified manner. Figure 1 shows the overview of DeepVS framework, which has three major components: the convolutional module, the attention module, and the multi-head regression module. We describe the details of these components in the following subsections.

3.1 1D CNN Module

The 1D CNN module is organized in a two-stream scheme to process the time domain input X^T and frequency domain input X^F in parallel. Recall that the input (as illustrated in Figure 1) is stacked with a sequence of consecutive time windows of signals in the time/frequency domain. To be specific, $X^T \in \mathbb{R}^{L \times M}$ and $X^F \in \mathbb{R}^{L \times N}$, where L is the number of consecutive time windows in a sequence, and M and N are the number of temporal and spectral samples in one time window, respectively. We use 1D CNN to capture mainly two kinds of local features, both of which indicate the existence of vital signs: the periodic variations in the time domain and the condensed spectral peak in the frequency domain.

3.2 Attention Module

We concatenate the output of the CNN module from time domain and frequency domain together to achieve the fused embedding ($X \in \mathbb{R}^{L \times d_{in}}$, $d_{in} = M + N$) as the input to the attention module.

In the attention module, we employ a self-attention mechanism from the encoder block of Transformer [16], which integrates the information within the entire sequence to capture temporal correlations across different time windows, in the form of a weighted sum, defined as:

$$\text{Attention}\ (Q, K, V) = \text{softmax}\left(\frac{QK^{\mathrm{T}}}{\sqrt{d_k}}\right)V, \tag{4}$$

where $Q = X\mathcal{W}_Q \in \mathbb{R}^{L \times d_k}$, $K = X\mathcal{W}_K \in \mathbb{R}^{L \times d_k}$, $V = X\mathcal{W}_V \in \mathbb{R}^{L \times d_V}$ are *queries*, *keys*, *values*, by projecting the input ($X \in \mathbb{R}^{L \times d_{in}}$) according to parameter matrices $\mathcal{W}_Q \in \mathbb{R}^{d_{in} \times d_k}$, $\mathcal{W}_K \in \mathbb{R}^{d_{in} \times d_k}$, and $\mathcal{W}_V \in \mathbb{R}^{d_{in} \times d_v}$. d_k is the dimension of *queries* and *keys*, and d_v is the dimension of *values*, while $\sqrt{d_k}$ is a scaling factor for smoother gradients. In implementation, we use 8-head attention [16].

3.3 Multi-head Regression Module

The output of the attention module is a series of vectors in a sequence of length L. We apply an average pooling to aggregate information of each vector and obtain the sequence level feature representation ($E \in \mathbb{R}^{L}$). It becomes the input to the multi-head regression module, which consists of two fully connected layers ($FC_i(E) = E\mathcal{W}_i + b_i$, $\mathcal{W}_i \in \mathbb{R}^{L \times 1}$ is weight matrix and b_i is a bias term), one for HR and the other one for RR in parallel. The end-to-end DeepVS is trained by optimizing the mean squared error against ground truth as the loss function propagated from two heads.

4 EVALUATION

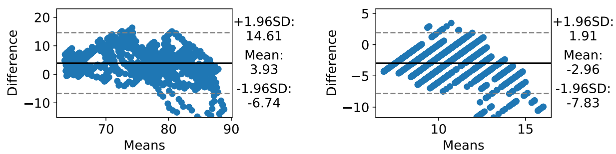
In this section, we first describe the data collection and datasets of RF-based vital signs sensing. Then we show the overall performance and ablation study.

4.1 Data Collection and Datasets

For RF-based vital signs data collection, we follow the same setting as [21] using a COTS UWB sensor (XeThru X4M03 [14]) as the RF front end, which samples the target chest wall displacement at 10 Hz. We obtain the ground truth of vital signs from a FDA approved medical device, Masimo Pulse Oximeter [1]. We use a sliding window of 30 seconds at 1 second increments and stack 30 consecutive time windows in a sequence as one data sample, thus the time domain input is of shape (30, 300) and frequency domain input (30, 150). We conduct data collection in home environments with 8 participants, following a pre-established protocol that protects the anonymity of the participants. We build two data sets. The first data set (denoted as $D1$) has in total $80,568$ samples for training. It was collected with 8 participants and each spent about 1 hour for data collection during sedentary behaviors (e.g., reading and typing) with intermittent body movements. Additionally, it includes vital signs data collected during sleep over two nights from one participant. The second data set (denoted as $D2$) is for testing independent from $D1$, and it has in total 1877 samples collected over 30 minutes during which participants have casual movements in sedentary behaviors, more challenging than settings in [20]. Note that, while both training ($D1$) and testing ($D2$) data sets include data collected during sedentary behaviors with intermittent body movements, they have different patterns (e.g., the duration and frequency of body movements), thus the testing data set will fail the overfitted model.

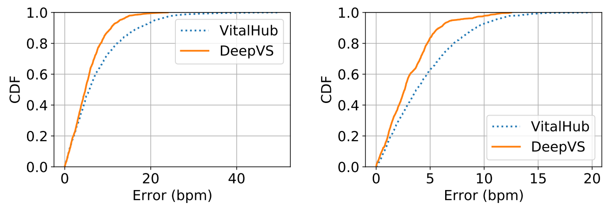
4.2 Overall Performance

We evaluate the overall performance of DeepVS on the data set $D2$, which is independent from the training data $D1$. Figure 2(a) and Figure 2(b) show the Bland-Altman plots of predicted HR and



(a) Bland–Altman plot of HR. (b) Bland–Altman plot of RR.

Figure 2: The Bland–Altman plots indicate the agreement between HR/RR and ground truth. The differences between prediction and ground truth of HR/RR are reasonably bounded within 95% limits of agreement (Mean$\pm 1.96SD$).



(a) CDF of HR. (b) CDF of RR.

Figure 3: CDF curves of the absolution errors show that DeepVS outperforms VitalHub for both HR and RR. Specifically, DeepVS has shorter tails than VitalHub, which indicate the better robustness achieved by DeepVS when dealing with challenging cases.

RR against the ground truth in data $D1$. Although they are not perfectly spreading flat indicating a perfect agreement, their 95% limits of agreement (mean±standard deviation of the difference) are bounded in a reasonable range.

We also compare the performance of DeepVS with a recent non-learning based SOTA solution VitalHub [20] (which were also designed to tackle the challenges from non-linearity and disturbance of body movements simultaneously) in terms of the absolution error against the ground truth. As shown in Figure 3(a) and Figure 3(b), DeepVS has shorter tails in the error distribution of both HR/RR than VitalHub. specifically, DeepVS reduces 80-percentile erros of HR/RR to 7.4/4.9 bpm from 11.8/7.3 bpm by VitalHub. It implies that DeepVS can better deal with challenges of non-linearity and body movements than VitalHub with heuristic-based algorithms.

4.3 Ablation Study

To demonstrate the effectiveness of DeepVS, we build 6 variants by referencing DeepVS framework (see Figure 1): a and b remove the time (T) and frequency (F) domain channels, respectively; c and d remove the CNN and attention modules, respectively; e and f have individual regression modules for HR and RR, respectively. Table 1 shows mean and standard deviation of absolution errors, the number of parameters and average inference latency of each setting.

Table 1: Ablation Study

Settings	HR err.	RR err.	#param	latency
a (w/o T)	5.67 ± 4.76	3.63 ± 2.47	148610	2.24 ms
b (w/o F)	5.68 ± 5.30	3.72 ± 2.13	153410	2.33 ms
c (w/o cnn)	5.86 ± 4.71	3.25 ± 2.09	154114	2.32 ms
d (w/o attn.)	6.15 ± 4.37	3.80 ± 2.23	**29026**	**1.84 ms**
e (HR only)	5.29 ± 4.62	-	166497	2.71 ms
f (RR only)	-	3.43 ± 2.09	166497	2.71 ms
DeepVS	**5.16** ± 3.82	**3.20** ± 2.35	166528	2.72 ms

We have three main observations from the results. 1) Combining inputs from time and frequency domain works better than using the individual input of either T or F. 2) Integrating local features and temporal correlations by combining the CNN and attention modules improves the accuracy. While the performance without the attention module is reasonably good with relatively small overhead, it has long-tail error from challenging cases. 3) Using the multi-head scheme with only negligible additional overhead, DeepVS has better estimation performance of both HR and RR than regression of HR or RR only, because it is optimized by both sources. In summary, DeepVS achieves the robustness of vital signs sensing by integrating different sources of information in a unified manner. In addition, the number of parameters and the inference latency shown in Table 1 indicate that the proposed model is efficient in terms of the space and time complexity of the model, promising for real-time deployment even in resource constrained edge devices.

5 DISCUSSION

As a preliminary exploration, DeepVS leaves a few directions to further studies. First, we only focus on the RF signal from a single user, while dealing with multi-user scenarios is more challenging. If the signal reflected from multiple users can be separated using spatial resolution with the given RF configurations, then the model works on the separated individual signals in the same way as on the single user. Otherwise, an extra step to achieve blind source separation will be necessary before using the current model for vital signs sensing. Second, we are yet to demonstrate the medical usefulness of DeepVS with potential applications such as sleep monitoring and health event detection. Finally, the generalizability of the model may be an issue when the data distribution of the target task differs a lot from the training data. A combination of heuristics and learning based solution may be considered to address this problem.

Code availability. We open source DeepVS: https://github.com/SBU-MoCA/DeepVS_bcb22.

6 CONCLUSION

We propose a hybrid framework by combining CNN and attention models to address key challenges from non-linearity and body movements for RF-based vital signs sensing in a unified manner. Experiments show that our proposed DeepVS largely reduce the 80-percentile error of HR/RR estimation compared to VitalHub from 11.8/7.3 bpm to 7.4/4.9 bpm.

REFERENCES

[1] [n. d.]. Masimo - MightySat Rx. https://www.masimo.com/products/monitors/spot-check/mightysatrx/
[2] Fadel Adib, Hongzi Mao, Zachary Kabelac, Dina Katabi, and Robert C Miller. 2015. Smart homes that monitor breathing and heart rate. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. ACM, 837–846.
[3] Zhe Chen, Tianyue Zheng, Chao Cai, and Jun Luo. 2021. MoVi-Fi: motion-robust vital signs waveform recovery via deep interpreted RF sensing. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*. 392–405.
[4] Pierre Comon and Christian Jutten. 2010. *Handbook of Blind Source Separation: Independent component analysis and applications*. Academic press.
[5] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. *Advances in neural information processing systems* 27 (2014).
[6] Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, Gang Wang, Jianfei Cai, et al. 2018. Recent advances in convolutional neural networks. *Pattern Recognition* 77 (2018), 354–377.
[7] Unsoo Ha, Salah Assana, and Fadel Adib. 2020. Contactless seismocardiography via deep learning radars. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*. 1–14.
[8] Peter Hillyard, Anh Luong, Alemayehu Solomon Abrar, Neal Patwari, Krishna Sundar, Robert Farney, Jason Burch, Christina Porucznik, and Sarah Pollard. 2018.

Experience: Cross-Technology Radio Respiratory Monitoring Performance Study. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*. ACM, 487–496.

[9] Aapo Hyvarinen and Hiroshi Morioka. 2016. Unsupervised feature extraction by time-contrastive learning and nonlinear ica. *Advances in Neural Information Processing Systems* 29 (2016).

[10] Changzhi Li and Jenshan Lin. 2008. Random body movement cancellation in Doppler radar vital sign detection. *IEEE Transactions on Microwave Theory and Techniques* 56, 12 (2008), 3143–3152.

[11] Feng Lin, Chen Song, Yan Zhuang, Wenyao Xu, Changzhi Li, and Kui Ren. 2017. Cardiac scan: A non-contact and continuous heart-based user authentication system. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*. 315–328.

[12] James C Lin. 1975. Noninvasive microwave measurement of respiration. *Proc. IEEE* 63, 10 (1975), 1530–1530.

[13] Jian Liu, Yan Wang, Yingying Chen, Jie Yang, Xu Chen, and Jerry Cheng. 2015. Tracking vital signs during sleep leveraging off-the-shelf wifi. In *Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing*. ACM, 267–276.

[14] Novelda LLC. [n. d.]. X4M03 Radar Development Kit. https://www.xethru.com/xethru-development-platform.html

[15] Van Nguyen, Abdul Q Javaid, and Mary Ann Weitnauer. 2014. Spectrum-averaged Harmonic Path (SHAPA) algorithm for non-contact vital sign monitoring with ultra-wideband (UWB) radar. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2241–2244.

[16] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).

[17] Xuyu Wang, Chao Yang, and Shiwen Mao. 2017. PhaseBeat: Exploiting CSI phase data for vital sign monitoring with commodity WiFi devices. In *ICDCS*.

[18] Zongxing Xie, Yindong Hua, and Fan Ye. 2022. A Measurement Study of FMCW Radar Configurations for Non-contact Vital Signs Monitoring. In *RadarConf22*.

[19] Zongxing Xie, Bing Zhou, Xi Cheng, Elinor Schoenfeld, and Fan Ye. 2021. Fusing UWB and Depth Sensors for Passive and Context-Aware Vital Signs Monitoring. In *IEEE/ACM CHASE 2021*.

[20] Zongxing Xie, Bing Zhou, Xi Cheng, Elinor Schoenfeld, and Fan Ye. 2021. VitalHub: Robust, Non-Touch Multi-User Vital Signs Monitoring using Depth Camera-Aided UWB. In *2021 IEEE 9th International Conference on Healthcare Informatics*.

[21] Zongxing Xie, Bing Zhou, and Fan Ye. 2021. Signal quality detection towards practical non-touch vital sign monitoring. In *ACM-BCB 2021*.

[22] Xiangyu Xu, Jiadi Yu, Yingying Chen, Yanmin Zhu, Linghe Kong, and Minglu Li. 2019. Breathlistener: Fine-grained breathing monitoring in driving environments utilizing acoustic signals. In *ACM Mobisys 2019*.