High Accuracy WiFi Sensing for Vital Sign Detection with Multi-Task Contrastive Learning

Yilun Wang*, Peng Cheng‡*, Shenghong Li†, Branka Vucetic*, and Yonghui Li*
*School of Electrical and Information Engineering, The University of Sydney, Australia
†Data 61, CSIRO, Australia
‡Department of Computer Science and Information Technology, La Trobe University, Australia
Email: Yilun.Wang@sydney.edu.au

***Abstract*—WiFi sensing has emerged as a promising technique in the healthcare industry, enabling contact-free monitoring of vital signs by detecting changes in WiFi signals resulting from physiological activities. State-of-the-art WiFi sensing uses channel state information (CSI) to analyze signal characteristics, capturing subtle changes due to heartbeats and breathing. However, existing methods face challenges in concurrently measuring respiration and heart rates, and they exhibit high sensitivity to environmental factors and individual differences, limiting the detection accuracy of a trained model in real-world environments. In this paper, we propose a novel multi-task contrastive learning framework for concurrent detection of respiration and heart rates. We introduce multi-task learning with hard-shared layers to exploit the physiological link between breathing and heartbeat. Additionally, we leverage contrastive learning to improve our model's ability to differentiate and prioritize CSI changes related to respiratory and cardiac activities. The experimental results demonstrate the proposed model's ability to accurately measure respiratory and heart rates in challenging scenarios, including long-distance and non-line-of-sight conditions, even when utilizing omnidirectional antennas.**

I. Introduction

WiFi sensing emerges as a transformative approach to environment perception [1], extending beyond its traditional role in communication. It leverages and harnesses the widespread availability of WiFi signals to detect subtle variations in signal characteristics, including magnitude, phase, and frequency caused by the presence of nearby objects and people. Common uses of WiFi sensing include locating and navigating people or objects indoors [2], identifying individuals for customized smart home experiences [3], and detecting postures for contact-free health monitoring [4].

WiFi is receiving increasing recognition for its great potential in the healthcare domain [5]. It enables contactless monitoring of vital signs by detecting subtle changes in WiFi signals from physiological activities like breathing and heartbeats. It is particularly beneficial for patients intolerable to traditional monitoring methods, such as those with burn injuries and elders living with dementia. Additionally, it allows for remote vital sign monitoring, which can decrease the need for frequent physical check-ups and minimize infection risks during post-operative recovery or infectious disease outbreaks.

Traditional WiFi-based vital sign detection depends on received signal strength (RSS), and thus suffers from low resolution, limited range, and vulnerability to ambient noise, resulting in inaccurate measurements [6]. The shift towards using channel state information (CSI) represents a substantial improvement. CSI provides more details about WiFi signal's characteristics, such as amplitude and phase across various frequency channels, thus enabling the precise capture of minute fluctuations caused by heartbeats and breathing. The authors in [7] successfully monitored heartbeats by employing a directional antenna to enhance the received signal quality and utilizing CSI phase data. However, relying solely on phase information may lead to a loss of details, potentially impacting detection accuracy. Additionally, the use of directional receiving antennas limits the versatility of the technology. Later work [8] integrated both amplitude and phase information to improve the performance of respiration and heartbeat tracking. Recent studies [9]–[11] have focused on integrating signal processing and machine learning to detect respiration and heartbeat, as well as monitor sleep stages by observing rate changes.

Despite extensive research in CSI-based vital sign monitoring, two primary challenges remain unaddressed. 1) Accurately measuring respiration and heart rate simultaneously is extremely challenging. The subtle heartbeat-induced variations in CSI are often masked by breathing movements, necessitating more sophisticated analysis of the CSI signals. 2) The system is highly sensitive to variations in both environmental conditions and individual characteristics, including age, gender, and breathing patterns. These variables can significantly reduce the detection accuracy when the trained model is deployed in diverse real-world environments.

In this paper, we propose a novel multi-task contrastive learning framework for concurrent detection of respiration and heart rates. To tackle the first issue, we introduce multi-task learning with hard-shared layers that exploit the inherent physiological link between breathing and heartbeat.

979-8-3503-8741-4/24/$31.00 ©2024 IEEE

We further develop task-specific Transformer architectures to provide distinct predictions for both respiration and heartbeat rates. This method not only achieves simultaneous measurements of these physiological signals but also avoids information loss typically associated with traditional signal separation methods. To overcome the second challenge, we leverage contrastive learning to enhance our model, enabling it to differentiate and prioritize the relevance of CSI changes to respiratory and cardiac activities, rather than treating all CSI-derived information equally. To mitigate the impact of environmental and individual differences on our model, we train the model with a comprehensive dataset and direct its attention towards relevant information to diminish its sensitivity to extraneous factors. Our experimental results show that with multi-task learning, the model yields accurate respiratory and heart rates, even with omnidirectional antennas, in various challenging scenarios such as long-distance and non-line-of-sight monitoring.

II. System and Model Design

In this section, we present the fundamental principle and system design of the proposed CSI-based vital sign monitoring scheme. We specify the data processing workflow, including pre-processing, data screening, and other related aspects.

A. CSI data

CSI reflects the environmental impact on WiFi signal propagation channel caused by nearby objects and people. Using modified drivers for a commercial WiFi adaptor, we can extract the frequency-domain CSI associated with each received WiFi frame. Considering a multi-path channel with L paths, with the i-th path characterized by a complex channel gain $A_i(t)$ and a path delay $d_i(t)$ at time t, the channel response in each sub-channel is given by

$$H(f,t) = \sum_{i=1}^{L} A_i(t) e^{-j2\pi f \frac{d_i(t)}{c}} + z(f,t), \tag{1}$$

where f denotes the carrier frequency of the sub-channel, c is the speed of light, and $z(f,t)$ is the measurement noise. Of the L paths, there are some static paths resulting from fixed environmental objects, while the remaining are dynamic paths with A_i and d_i varying with various movements in the environment, including the chest movement due to breathing and heartbeat. Accordingly, the channel response $H(f,t)$ can be re-written as

$$H(f,t) = H_s(f,t) + H_d^e(f,t) + H_d^c(f,t) + z(f,t) \tag{2}$$

where $H_s(f,t)$, $H_d^e(f,t)$, and $H_d^c(f,t)$ denote the static component, the dynamic component due to background movements, and the dynamic component due to the chest movement of the individual under monitoring. Due to the interference of $H_s(f,t)$, $H_d^e(f,t)$, and $z(f,t)$, sophisticated data processing algorithm is required to extract the breathing and heartbeat rates from $H(f,t)$.

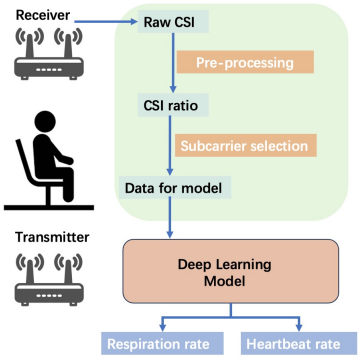


Fig. 1: The system design of vital sign detection.

B. Data Processing Flow

Fig. 1 illustrates the data processing workflow of the proposed system. Raw CSI data first undergoes a two-step preprocessing routine for noise and interference elimination. The first step involves compensating for random excess delay caused by the lack of time synchronization between the WiFi transmitter and receiver. The second step merges the amplitude and phase of the CSI, followed by subcarrier selection. The processed data is then input into the proposed multi-task contrastive learning model that predicts breaths and heartbeats per minute (bpm). A clinical-grade contact-based sensor is used to collect the ground-truth data for training the model.

C. Data Pre-processing and Subcarrier Selection

Due to randomness in the triggering circuitry of WiFi receivers, each received WiFi frame is subject to an unknown excess delay $\varphi(t)$, which causes a linear change across sub-channels in the phase of CSI. Specifically, the received CSI, affected by random time offsets, is given by

$$\tilde{H}(f,t) = e^{-j2\pi f \frac{\phi(t)}{c}} H(f,t). \tag{3}$$

If not addressed, the phase of $\tilde{H}(f,t)$ changes randomly over time, making it even harder to recover useful information.

It is noted that antennas on the same WiFi adaptor use a shared clock, resulting in consistent $\varphi(t)$ across them. By dividing one antenna's CSI by the other's, we can eliminate the random excess delay while preserving the environmental impact on CSI [12]. Specifically, we calculate a ratio between the CSIs collected on two receive antennas as below

$$\rho = \frac{\tilde{H}_1(f,t)}{\tilde{H}_2(f,t)} = \frac{H_1(f,t)}{H_2(f,t)} \tag{4}$$

where $\tilde{H}_1(f,t)$ and $\tilde{H}_2(f,t)$ represent the CSIs from the first and second receive antennas, respectively.

After reducing the randomly excess delay, the CSI is processed through a Savitzky-Golay filter. The filter effectively reduces the impact of noise $z(f,t)$ while maintaining the influence of breathing and heartbeat on CSI. Following the method from [13], we project both real $\mathcal{R}(\rho)$ and imaginary parts $\mathcal{I}(\rho)$ of ρ onto different angles in the complex plane, generating a new signal to represent changes in CSI. The projection $\mathcal{P}$ is shown as

$$\mathcal{P} = [cos\theta \ sin\theta] \left[\mathcal{R}(\rho) \ \mathcal{I}(\rho)\right]^T. \tag{5}$$

We divide the range of θ, from 0 to 2π, into 200 steps (candidates) of $\pi/50$ each, generating 200 candidates for ρ. This method ensures dynamically adjusts their orthogonal components. As a result, if either $\mathcal{R}(\rho)$ or $\mathcal{I}(\rho)$ is unsuitable for vital sign detection at any point, the other will gain more weight in the combination. We identify the optimal angle by comparing energy ratios. Setting the bin limits as the lowest normal respiration rate (10 bpm) and highest normal heart rate (120 bpm), we perform FFT on all candidates and calculate an energy ratio within this frequency range against total spectral energy. The candidate with the highest ratio represents vital signs best, as it exhibits minimal noise interference from $z(f,t)$ and is least significantly influenced by $H_d^c(f,t)$.

Multi-path effects result in varying channel conditions for different subcarriers, leading to diverse levels of attenuation and phase shifts. Some subcarriers remain unaffected by breathing and heartbeat, necessitating their filtration and the selection of those reflecting these influences. Preliminary experiments show that the hardware we utilized typically has 5 to 7 subcarriers that are clearly influenced by breathing and heartbeat following pre-processing. Consequently, the data fed into the deep learning model comprises 5 subcarriers. The method for selecting subcarriers is also based on energy ratios. In each individual subcarrier, we select only the optimal angle. This angle's energy ratio represents the energy ratio of that subcarrier. During the process of selecting subcarriers, we will compare the energy ratios of different subcarriers and choose the best five. The entire process is summarized in Algorithm 1.

Algorithm 1 Select Subcarrier With Best Projection Angle

function SUBCARRIER SELECTION($data, \theta, fft\_size$)
  $subcarrier\_results \leftarrow \emptyset$
  for each $subcarrier$ in $data$ **do**
    $angle\_list \leftarrow \emptyset$
    for each θ_i in θ **do**
      $projection = [\cos(\theta_i), \sin(\theta_i)] \cdot$ [Real($subcarrier$), Imag($subcarrier$)]^T
      Apply zero-padding $\rightarrow fft\_size$
      $vital\_energy \leftarrow$ Energy($[10/60, 120/60]$ Hz)
      $full\_energy \leftarrow$ Energy(full spectrum)
      $energy\_rate \leftarrow vital\_energy / full\_energy$
      Add $(energy\_rate, projection)$ to $angle\_list$
    end for
    Sort $angle\_list$ by $energy\_rate$ descending
    Get $(max\_rate, best\_proj)$ from $angle\_list$
    Add $(max\_rate, best\_proj)$ to $subcarrier\_results$
  end for
  Sort $subcarrier\_results$ by $energy\_rate$ descending
  $final\_results \leftarrow$ Top 5 projections in $subcarrier\_results$
  return $final\_results$
end function

III. Multi-Task Contrastive Learning

In this section, we will develop a novel deep learning model that integrates multi-task learning with contrastive learning. The model performs two primary tasks: obtaining the BPM values of breathing and heartbeat. Additionally, a third hidden task is designed to enhance the model's accuracy under different environments and test individuals through contrastive learning.

A. Overview

The model architecture is depicted in Fig. 2. The model takes two-dimensional pre-processed CSI data as input, which includes data from 5 selected subcarriers, all within the same 60-second window (T = 6000, sampling rate = 100Hz). The model's output provides the respective bpm values for respiration and heartbeat.

This model's fundamental structure is primarily divided into two main components. The first part is comprised of hard-shared layers, including three CNN layers ($\mathbf{f}(\mathbf{x})$). Each of these layers is equipped with batch normalization and is followed by a dropout layer to prevent over-fitting. These layers, with parameters $\mathbf{\Theta}_1$, are designed to extract respiratory and cardiac features from the input CSI data. The extracted features are then processed further. In the context of contrastive learning, the model is trained to focus more on extracting information specific to breathing and heartbeat. This is achieved by designing the loss function to penalize the model more heavily if it fails to ignore variations caused by different environments and individuals. Consequently, this approach facilitates the refinement and update of the parameters $\mathbf{\Theta}_1$, enhancing the model's ability to extract relevant features under varied conditions. The second part comprises two Transformer regression models with parameters $\mathbf{\Theta}_2$ and $\mathbf{\Theta}_3$. Here, an attention-based custom encoder encodes the anchor and positive features set, which then passes through two linear layers to determine respiration and heartbeat rates. The model computes three losses: $\mathcal{L}_1$ and $\mathcal{L}_2$ from the regression models, and contrastive learning-related loss $\mathcal{L}_3$. To prevent overfitting, early stopping will be triggered when all losses do not show improvement within 10 epochs, and the best $\mathbf{\Theta}_1$, $\mathbf{\Theta}_2$, and $\mathbf{\Theta}_3$, will be saved.

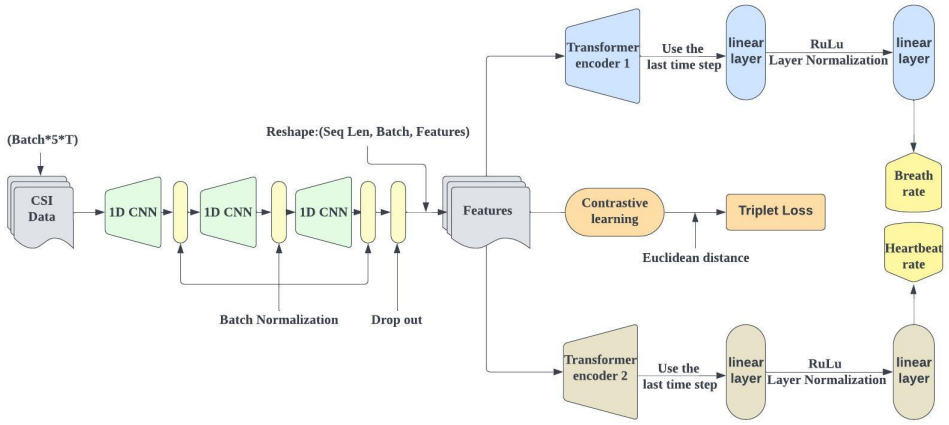


Fig. 2: The proposed multi-task contrastive learning model.

B. Loss function Design

We derive $\mathcal{L}_1$ and $\mathcal{L}_2$ directly from the differences between predicted and label values (ground truth) in the regression model, which effectively measures the model's performance. We employ mean square error (MSE) to calculate the difference in bpm values, generating $\mathcal{L}_1$ and $\mathcal{L}_2$.

To calculate $\mathcal{L}_3$, we first determine the Euclidean distance between the features of positive/ negative and anchor samples, which is given by

$$D(\mathbf{X}_A, \mathbf{X}_{P/N}) = \left\| \mathbf{f}(\mathbf{X}_A) - \mathbf{f}(\mathbf{X}_{P/N}) \right\|_2, \tag{6}$$

where $\mathbf{X}_A$ denotes anchor CSI data and $\mathbf{X}_{P/N}$ represents either positive and negative CSI data, respectively.

The distance $D(\mathbf{X}_A, \mathbf{X}_{P/N})$, once calculated, are used in back-propagation of $\mathcal{L}_3$ through Triplet Loss [14]. When the condition $D(\mathbf{X}_A, \mathbf{X}_P) + \alpha \leq D(\mathbf{X}_A, \mathbf{X}_N)$ is satisfied, Triplet Loss ensures stability by setting both $\mathcal{L}_3$ and all related gradients to zero. This method effectively prevents over-fitting and preserves crucial features. The calculations for $\mathcal{L}_3$ can be shown as follows.

$$\mathcal{L}_3 = \max(D(\mathbf{X}_A, \mathbf{X}_P) - D(\mathbf{X}_A, \mathbf{X}_N) + \alpha, 0). \tag{7}$$

Here, α is a margin that is added to the distance between the anchor and the positive sample. It acts as a threshold to ensure that the positive and negative samples are separated by a certain distance.

C. Multi-loss Optimization

In multi-task learning, managing gradient sets effectively is key to optimizing model performance. In this paper, we introduce a method that concentrates on gradient management to ensure the final convergence of the model. Specifically, for the two Transformer structures in the model, updating $\mathbf{\Theta}_2$ and $\mathbf{\Theta}_3$ is relatively straightforward. It only requires gradient calculation based on the corresponding losses $\mathcal{L}_1$ and $\mathcal{L}_2$, followed by backpropagation. Due to different convergence speeds, different optimizers are designed for the two tasks to adjust the learning rates (η_2, η_3). The process of gradient calculation and parameter updating is as follows.

$$\begin{aligned} \mathbf{G}_{\text{Transformer1}} &= \nabla_{\mathbf{\Theta}_2} \mathcal{L}_1, \\ \mathbf{G}_{\text{Transformer2}} &= \nabla_{\mathbf{\Theta}_3} \mathcal{L}_2, \end{aligned} \tag{8}$$

$$\begin{aligned} \mathbf{\Theta}_2 &= \mathbf{\Theta}_2 - \eta_2 \mathbf{G}_{\text{Transformer1}}, \\ \mathbf{\Theta}_3 &= \mathbf{\Theta}_3 - \eta_3 \mathbf{G}_{\text{Transformer2}}, \end{aligned} \tag{9}$$

where $\mathbf{G}_{\text{Transformer1}}$ and $\mathbf{G}_{\text{Transformer2}}$ respectively represent the gradients of the loss function $\mathcal{L}_1$ with respect to the parameters $\mathbf{\Theta}_2$ and the gradients of the loss function $\mathcal{L}_2$ with respect to the parameters $\mathbf{\Theta}_3$.

Updating $\mathbf{\Theta}_1$ is complex as it requires adjustments from $\mathcal{L}_1$, $\mathcal{L}_2$, and $\mathcal{L}_3$. Given their varying magnitudes, each loss should be weighted and assigned a separate optimizer with a specific learning rate η_1. On this basis, we have

$$\begin{aligned} \mathbf{G}_{\text{CNN}} &= \alpha_1 \nabla_{\mathbf{\Theta}_1} \tilde{\mathcal{L}}_1 + \alpha_2 \nabla_{\mathbf{\Theta}_1} \tilde{\mathcal{L}}_2 + \alpha_3 \nabla_{\mathbf{\Theta}_1} \tilde{\mathcal{L}}_3, \\ \mathbf{\Theta}_1 &= \mathbf{\Theta}_1 - \eta_1 \mathbf{G}_{\text{CNN}}, \end{aligned} \tag{10}$$

where α_1, α_2, and α_3 represent the contribution of each of the three types of loss during each gradient update. For the calculation of these contributions, dynamic weight adjustment methods [15] are typically used, which update the loss weight in each iteration based on historical performance. In this method, tasks with lower performance are assigned higher weights during the next update. The weight for task k is calculated as follows

$$\begin{aligned} \alpha_k(t) &= \frac{K e^{(w_k(t-1)}/T)}{\sum_i e^{(w_k(t-2)/T)}}, \\ w_k(t-1) &= \frac{\mathcal{L}_k(t-1)}{\mathcal{L}_k(t-2)}, \end{aligned} \tag{11}$$

where K and T are constants used to adjust sensitivity, and $\sum_i e^{(w_i(t-1)/T)}$ guarantees the total sum of weights for all tasks is 1. The term $w_k(t-1)$ represents the historical performance.

In our multi-task model, each task is a convex optimization problem with non-conflicting gradients ($\mathbf{G}_k = \nabla_{\mathbf{\Theta}} \mathcal{L}_k$), satisfying $\cos(\mathbf{G}_i, \mathbf{G}_j) \geq 0$ for all i, j. Therefore, assigning individual optimizers to each task ensures global optimality without the need for considering task weights [16]. The overall loss function is given by

$$\min_{\mathbf{\Theta}} \mathcal{L}^{MT}(\mathbf{\Theta}) = \sum_{k \in \mathcal{T}} \mathcal{L}_k(f(\mathbf{\Theta}_k, \mathcal{X}, k), Y), \qquad (12)$$

where $\mathcal{T}$ represents the set of tasks and Y denotes the true labels. The prediction for task k with input $\mathcal{X}$, based on parameters $\mathbf{\Theta}_k$, is denoted by $f(\mathbf{\Theta}_k, \mathcal{X}, k)$. In a hard-sharing setup, the direct summation of losses and the subsequent back-propagation forces the model to find a balance point that simultaneously satisfies the needs of all tasks, rather than maximizing the performance of a single task. This acts as a mean constraint. It also compels the model to learn more general and representative features rather than those specific to a single task, thereby preventing over-fitting [17].

IV. Experimental Results

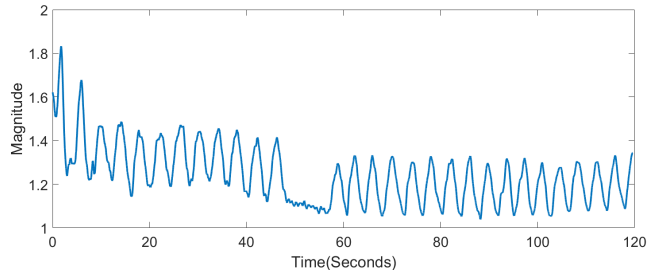
A. Experiment Setup

We collected data using Intel 5300 and CSI Tool [18] at a frequency of 100 Hz with a 3x3 antenna setup. Respiratory and heart rates were recorded as labels at 10 Hz using NUL-208 and NUL-236 devices. The data set, comprising about 300,000 entries from 14 anonymized volunteers (9 males, 5 females aged between 20-28) in an indoor setting, was divided into training (80%), validation (20%), and testing sets (3000 entries). We have set three scenarios: simple, normal, and challenging. In the simple scenario, the subject sits normally between the transmitter and receiver, 1 meter from each. In the normal scenario, the distance varies between 2-5 meters. The challenging scenario includes sitting sideways, distances over 6 meters, and non-line-of-sight conditions. We treat data from simple scenario as anchor samples and data from normal and challenging scenarios as positive samples. Negative samples are composed of CSI data that includes people moving and significant body movements. We developed the model on AWS P3.2xlarge servers using PyTorch. An illustration of the experiment environment is shown in Fig. 3.

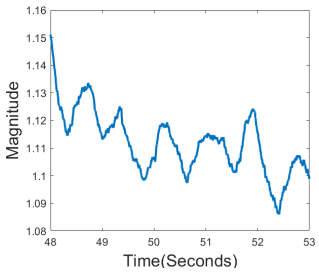
B. Effectiveness of Vital Sign Detection

With the correct hardware configuration and optimized data processing, breathing and heartbeat effects on CSI can be clearly observed in our indoor environment. Fig. 4(a) shows the regular variations of a subcarrier signal corresponding to breathing, including a breath-hold interval. A closer look

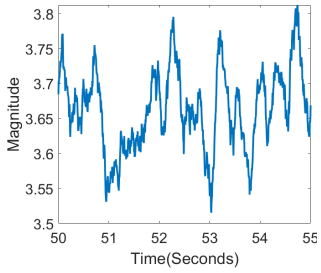
Fig. 3: An illustration of the experiment environment.



(a) Full Time Data



(b) Heartbeat Only Data



(c) Noise Data

Fig. 4: Effectiveness of the proposed system model in vital sign detection.

a this interval (Fig. 4(b)), reveals periodic CSI fluctuations due to heartbeats, clearly distinguishable from environmental noise (Fig. 4(c)). This confirms that CSI can detect chest movements caused by respiration and heartbeats. However, it is worth noting that during normal breathing, heartbeat-induced changes in CSI tend to be obscured.

C. Performance of the Proposed Model

We first evaluate the multi-task learning model using data from the simple scenario with a single volunteer. Fig. 5 compares our model to existing methods, including traditional and machine learning methods [7], [19]–[21], focusing on the comparison of average errors. Our multi-task model, compared to separate handling, decreases the average breathing

error by 32% and heartbeat error by 37%. Fig. 6 shows the cumulative distribution function (CDF) detection error performance. It is observed that 80% of respiratory rate errors are below 0.26 bpm and 80% of heart rate errors are below 1.21 bpm. In contrast, the separate handling approach records 80% of respiratory and heart rate errors below 0.62 bpm and 2.52 bpm, respectively. At the same time, our model achieves a maximum heart rate error of 3 bpm, which is only half of the traditional methods.

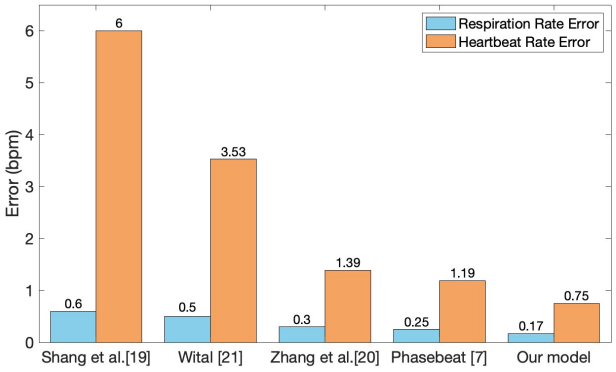


Fig. 5: Comparison with other methods.

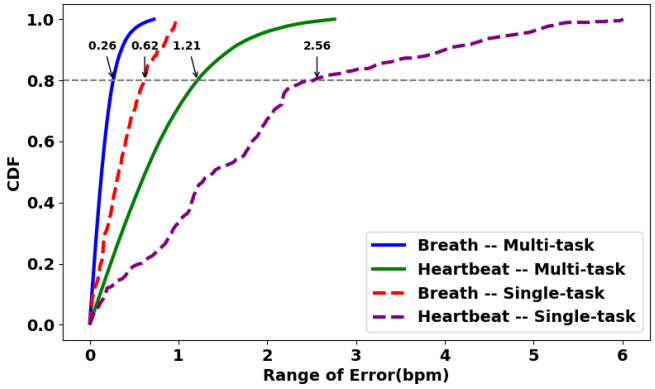


Fig. 6: Multi-task learning performance.

The detection accuracy of multi-task learning decreases when applied to normal and challenging scenarios and more volunteers. To solve this issue, we leverage contrastive learning, which maintains low error levels. Fig. 7 shows that the average errors for respiratory and heart rates were 0.27 bpm and 1.03 bpm, respectively, with 80% of errors under 0.58 bpm (respiration) and 2.03 bpm (heart rate). Compared to the results before incorporating contrastive learning, the average error in respiration decreased by 60%, while the average error in heartbeat decreased by 51%. This indicates that contrastive learning effectively targets CSI changes related to breathing and heartbeats over a scene or individual variations. The model also achieves high accuracy in completely new environments and with new volunteers, maintaining an average heart rate error of around 2 bpm. Our multi-task contrastive

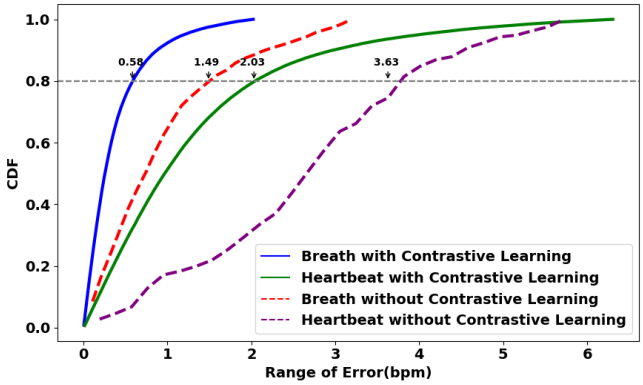


Fig. 7: Contrastive learning performance.

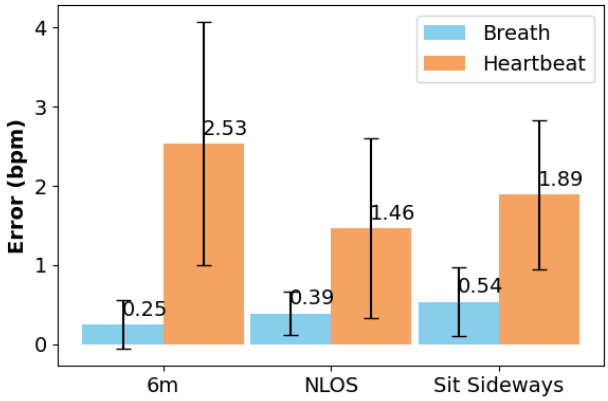


Fig. 8: Detection errors in challenging scenarios.

learning model can also achieve favourable performance in challenging scenarios with signal strength reduction, ample noise, or changes in the reflection path. At a distance of 6 meters, the average errors are 0.25 bpm for breathing and 2.53 bpm for heart rate; in non-line-of-sight conditions, these figures were 0.39 bpm and 1.46 bpm, respectively; when the subject is sitting sideways, they increase to 0.54 bpm and 1.89 bpm respectively. Given the bandwidth and transmission power limitations of WiFi sensing, this slight decrease in precision is still within an acceptable error range.

From Figs. 5-8, we can conclude that our multi-task contrastive learning model improves prediction accuracy by simultaneously managing the impacts of respiration and heartbeat on CSI. This model also adapts to scenario and subject variations, maintaining high accuracy even in challenging environments.

V. Conclusion

This paper introduced a novel multi-task constrastive learning framework for CSI-based vital sign detection. It employs a hard-shared layer to correlate the impacts of breathing and heartbeats on CSI, while simultaneously determining their BPM values. By integrating contrastive learning, the approach differentiates the changes in CSI, minimizing the influence of environmental and individual variations and focusing on alterations induced by breathing and heartbeats. Based on the collected experimental data, multi-task learning achieved a significant reduction in mean errors for respiration and heartbeat by 32% and 37%, respectively, when compared to frequency-based signal separation methods. This resulted in errors of 0.17 bpm and 0.75 bpm. Additionally, the integration of contrastive learning has reduced the average errors for respiration and heartbeat by about 60% and widened its applicability. It can maintain heartbeat errors at around 2 bpm even with new individuals and environments, reducing training costs and improving model generalization.

References

[1] Y. Ma, G. Zhou, and S. Wang, "Wifi sensing with channel state information: A survey," *ACM Computing Surveys (CSUR)*, vol. 52, no. 3, pp. 1–36, Jun. 2019.

[2] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti, "Spotfi: Decimeter level localization using wifi," in *Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication*, ser. SIGCOMM '15. New York, NY, USA: Association for Computing Machinery, Aug. 2015, p. 269–282. [Online]. Available: https://doi.org/10.1145/2785956.2787487

[3] J. Zhang, B. Wei, W. Hu, and S. S. Kanhere, "Wifi-id: Human identification using wifi signal," in *2016 International Conference on Distributed Computing in Sensor Systems (DCOSS)*, May. 2016, pp. 75–82.

[4] H. Abdelnasser, M. Youssef, and K. A. Harras, "Wigest: A ubiquitous wifi-based gesture recognition system," in *2015 IEEE Conference on Computer Communications (INFOCOM)*, Aug. 2015, pp. 1472–1480.

[5] D. Wu, Y. Zeng, F. Zhang, and D. Zhang, "Wifi csi-based device-free sensing: from fresnel zone model to csi-ratio model," *CCF Trans. on Pervasive Comp. and Interact.*, vol. 4, pp. 88–102, Sep. 2021.

[6] N. Patwari, L. Brewer, Q. Tate, O. Kaltiokallio, and M. Bocca, "Breathfinding: A wireless network that monitors and locates breathing in a home," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 1, pp. 30–42, Feb. 2013.

[7] X. Wang, C. Yang, and S. Mao, "Phasebeat: Exploiting csi phase data for vital sign monitoring with commodity wifi devices," in *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*, Jun. 2017, pp. 1230–1239.

[8] Y. Zeng, D. Wu, R. Gao, T. Gu, and D. Zhang, "Fullbreathe: Full human respiration detection exploiting complementarity of csi phase and amplitude of wifi signals," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, no. 3, Sep. 2018. [Online]. Available: https://doi.org/10.1145/3264958

[9] M. Liu, Z. Lin, P. Xiao, and W. Xiang, "Human biometric signals monitoring based on wifi channel state information using deep learning," *arXiv preprint arXiv:2203.03980*, Mar. 2022.

[10] M. I. Khan, M. A. Jan, Y. Muhammad, D.-T. Do, A. u. Rehman, C. X. Mavromoustakis, and E. Pallis, "Tracking vital signs of a patient using channel state information and machine learning for a smart healthcare system," *Neural Comput. and Applic.*, pp. 1–15, Jan. 2021.

[11] B. Yu, Y. Wang, K. Niu, Y. Zeng, T. Gu, L. Wang, C. Guan, and D. Zhang, "Wifi-sleep: Sleep stage monitoring using commodity wi-fi devices," *IEEE Internet of Things Journal*, vol. 8, no. 18, pp. 13 900–13 913, Mar. 2021.

[12] Y. Zeng, D. Wu, J. Xiong, E. Yi, R. Gao, and D. Zhang, "Farsense: Pushing the range limit of wifi-based respiration sensing with csi ratio of two antennas," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 3, no. 3, Sep. 2019. [Online]. Available: https://doi.org/10.1145/3351279

[13] Y. Zeng, D. Wu, J. Xiong, J. Liu, Z. Liu, and D. Zhang, "Multisense: Enabling multi-person respiration sensing with commodity wifi," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 4, no. 3, Sep 2020. [Online]. Available: https://doi.org/10.1145/3411816

[14] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 815–823, Jun 2015.

[15] S. Liu, E. Johns, and A. J. Davison, "End-to-end multi-task learning with attention," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, Jun. 2019, pp. 1871–1880. [Online]. Available: https://doi.ieeecomputersociety.org/10.1109/CVPR.2019.00197

[16] V. Kurin, A. De Palma, I. Kostrikov, S. Whiteson, and P. K. Mudigonda, "In defense of the unitary scalarization for deep multi-task learning," *Advances in Neural Information Processing Systems*, vol. 35, pp. 12 169–12 183, Nov. 2022.

[17] S. Ruder, "An overview of multi-task learning in deep neural networks," *CoRR*, vol. abs/1706.05098, 2017. [Online]. Available: http://arxiv.org/abs/1706.05098

[18] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11n traces with channel state information," *SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 1, p. 53, Jan. 2011.

[19] J. Shang and J. Wu, "Fine-grained vital signs estimation using commercial wi-fi devices," in *Proceedings of the Eighth Wireless of the Students, by the Students, and for the Students Workshop*, ser. S3. New York, NY, USA: Association for Computing Machinery, Oct. 2016, p. 30–32. [Online]. Available: https://doi.org/10.1145/2987354.2987360

[20] R. Zhang, Z. Wang, G. Li, Y. Wang, J. Shuai, J. Zheng, J. Huang, and S. Bao, "Hybrid subcarrier selection method for vital sign monitoring with long-term and short-term data considerations," *IEEE Sensors Journal*, vol. 22, no. 23, pp. 23 209–23 220, Oct. 2022.

[21] X. Zhang, Y. Gu, H. Yan, Y. Wang, M. Dong, K. Ota, F. Ren, and Y. Ji, "Wital: A cots wifi devices based vital signs monitoring system using nlos sensing model," *IEEE Transactions on Human-Machine Systems*, vol. 53, no. 3, pp. 629–641, Apr. 2023.