SeatDot: Semi-supervised Representation Learning for Contactless Vital Signs Estimation

Yingjian Song, Zaid Farooq Pitafi, Bingnan Li,
Yuan Ke, Bradley G. Phillips, Susan Brasher, Gari Clifford and Wenzhan Song

***Abstract*—In this paper, we introduce SeatDot, a contactless system for continuous vital signs monitoring on seat. SeatDot operates without user interaction and is easily installed under a seat. Unlike computer vision with abundant public datasets, gathering labeled data for sensor-based deep learning is challenging. SeatDot employs self-supervised contrastive learning, allowing us to extract robust feature representations from unlabeled data, which is more practical to gather than extensive labeled data. Additionally, SeatDot incorporates supervised representation learning, which aligns the differences between pairs of representations with the differences in their associated labels and requiring only 5% and 15% labeled data on HR and RR estimation, respectively. SeatDot's strength lies in its ability to leverage information from both labeled and unlabeled samples, enhancing model generalization and overall accuracy. Experimental results shows SeatDot achieves MAE of 2.44 BPM for heart rate (HR) and 2.20 BPM for respiration rate (RR) estimation, compare against an FDA-approved device. Rigorous testing on data from 32 subjects highlights SeatDot's reliability and potential for healthcare and wellness tracking, especially in smart office settings.**

***Index Terms*—semi-supervised representation learning, contactless vital signs, contrastive learning, heart rate, respiratory rate**

I. Introduction

Contactless vital signs estimation has emerged as a groundbreaking technology with vast applications in healthcare, wellness monitoring, and smart office environments [1]–[5]. The capability to accurately and reliably measure vital signs, such as heart rate and respiratory rate, without any physical contact, holds significant promise in elevating patient care and advancing well-being.

Yingjian Song, Zaid Farooq Pitafi and Wenzhan Song are with the School of Electrical and Computer Engineering, University of Georgia, Athens, GA 30602 USA. (e-mail: yingjian.song@uga.edu, zp18941@uga.edu, wsong@uga.edu). Wenzhan's research is partially supported by NSF-STTR-1940864, Georgia Research Alliance, DOE-EE0009026, NSF-SaTC-2019311, DOD-FA8571-21-C-0020.

Bingnan Li and Yuan Ke are with Department of Statistics, University of Georgia, Athens, GA 30602 USA. (e-mail: bingnan.li1@uga.edu, yuan.ke@uga.edu). Yuan's research is partially supported by NSF-DMS-2210468.

Bradley G. Phillips is with Biomedical And Translational Sciences Institute, University of Georgia, Athens, GA 30602 USA. (e-mail: bgp@uga.edu). Bradley's research is partially supported by Georgia CTSA UL1TR002378, NSF-1940864 and NSF-2302890.

Susan Brasher is with the School of Nursing, Emory University, Atlanta, GA 30322. (e-mail: susan.n.brasher@emory.edu)

Gari Clifford is with the Department of Biomedical Informatics, Emory University, Atlanta, GA 30322. (e-mail: gclifford6@gatech.edu)

Acknowledgement: thank Stephanie Croyle, Kimberly Schmitz and all UGA CTRU staff for helping the IRB and data collection experiments.

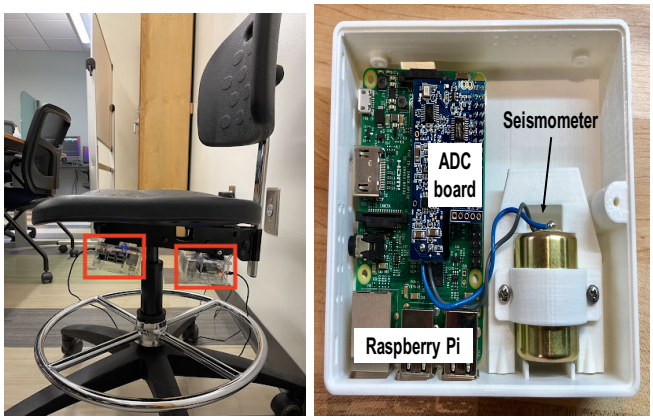

Fig. 1. Installation of SeatDot (in red rectangle) under the chair(left), SeatDot device(Right).

Various methods for vital sign monitoring have been explored in the literature. *Wearables* [6]–[8]: These devices have gained popularity due to their convenience and portability. However, wearables necessitate direct skin contact, which can be uncomfortable for some individuals. Additionally, users often overlook recharging them, and maintaining consistent patient engagement proves challenging. *Pads*: This category of monitoring devices features pads equipped with Micro-Electro-Mechanical Systems (MEMS), Field-Programmable Gate Array (FPGA), and pressure sensors [1]–[3]. Typically positioned between a mattress and its box spring or beneath the bedsheet, these pads are not ideal for seated use cases. *Side Monitors*: This category encompasses radar systems [4], [5] and camera-based solutions [9], [10]. While radars measure vital signs by emitting probing waves towards the heart and interpreting the reflections, they bring forth concerns about continuous radiation exposure and necessitate precise setup in terms of orientation and distance. On the other hand, camera-based solutions, especially those utilizing thermal cameras, have privacy implications, especially when set up in private spaces like bedrooms. Proper deployment of these technologies usually demands careful adjustments to effectively capture the subject's face. It's notable that the realm of contactless vital sign estimation in seated environments remains under-researched, especially in non-camera-based methodologies.

In this paper, we introduce SeatDot, an innovative framework tailored for enhancing the accuracy and robustness of contactless vital signs estimation in seated scenarios using seismic sensors. SeatDot effectively addresses the aforemen-

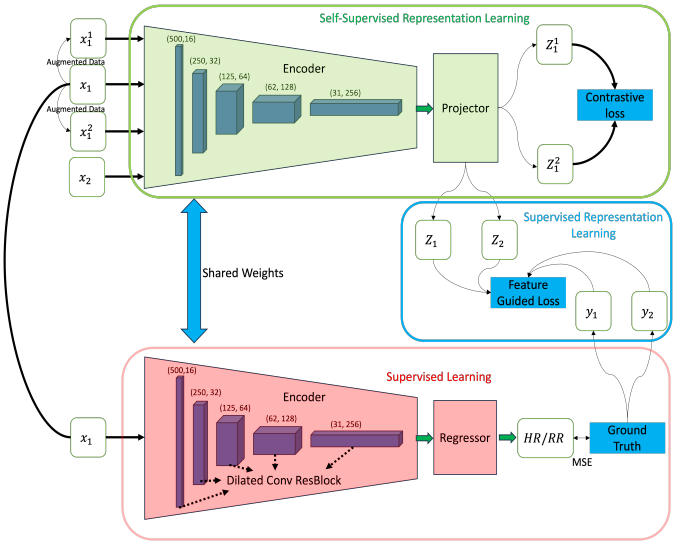

Fig. 2. SeatDot Training Framework.

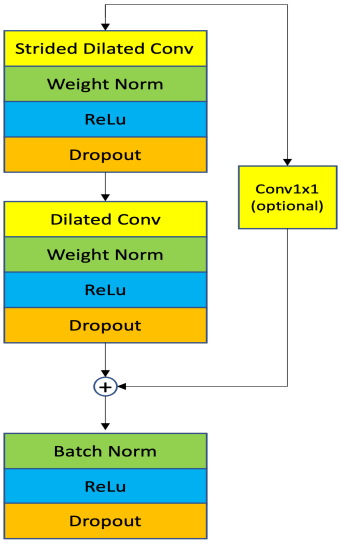

Fig. 3. Temporal Residual Block

tioned challenges and offers a non-invasive, practical approach to capturing important physiological data that can be implemented in clinical, research, and educational settings. While SeatDot can be used across the lifespan, it holds tremendous potential for the unobtrusive measurement and monitoring of vital signs in specific populations where this type of data can be more difficult to collect. Examples of populations include children with anxiety, attention deficit hyperactivity disorder (ADHD), Posttraumatic Stress Disorder (PTSD), medical trauma, and autism spectrum disorder (ASD), often face challenges when it comes to using wearable devices for data collection. In these scenarios, SeatDot offers a practical and valuable alternative. It prioritizes privacy by not capturing images or audio, obviates the need for direct physical contact with the human body, and sidesteps the necessity for active patient participation. SeatDot's installation is straightforward and hassle-free, as illustrated in Fig. 1. The core innovation of SeatDot lies in its integration of both supervised representation learning and unsupervised contrastive representation learning. This combination yields feature representations that are not only deeply informative but also unique, thereby bolstering the overall precision of vital sign estimation.

Deep learning model typically relies on large datasets for training, a process often hampered by the daunting task of comprehensive labeling. SeatDot navigates this hurdle by harnessing both a large dataset devoid of labeled annotations and a more modest set of labeled data. SeatDot addresses this challenge by leveraging a combination of a vast dataset without labeled annotations and a smaller set of labeled data. This dual approach taps into self-supervised contrastive learning, unraveling latent information within the data and formulating feature representations that capture intricate patterns and interrelations among vital signs. Moreover, SeatDot incorporates a feature-guided loss function that synchronizes the learned representations with the actual vital sign labels. This feature-guided loss function enables SeatDot to effectively map the differences between pairs of feature representations to the corresponding differences in vital sign labels, contributing to the model's overall accuracy. Our experimental results underscore the effectiveness of SeatDot, demonstrating its remarkable performance with a Mean Absolute Error (MAE) of 2.44 BPM for heart rate (HR) estimation and 2.20 BPM for respiratory rate (RR) estimation when compared to an FDA-approved device. This shows the potential of SeatDot to transform healthcare monitoring and wellness tracking, particularly in smart office environments.

We summarized our contributions as follows:

1) We designed and demonstrated an engagement-free vital signs monitoring system for seated scenarios, capable of real-time, accurate estimation of both heart rate and respiratory rate.
2) We formulate a two-step approach to enhance HR and RR estimation. First, self-supervised contrastive representation learning extracts valuable information from vast unlabeled data. Then, supervised representation learning aligns the information with actual vital sign labels. Notably, this approach achieves exemplary performance without necessitating the collection of an extensive volume of labeled data.
3) We conducted a validation test using a dataset collected from 32 patients, with the results showcasing SeatDot's commendable performance. SeatDot achieved a Mean Absolute Error (MAE) of 2.44 BPM for HR estimation and 2.20 BPM for RR estimation when compare against an FDA-approved device. Importantly, there is no patient overlap nor information leakage between training and testing datasets. This shows the model's robustness and its ability to generalize effectively across different users.

II. Related Works

In existing studies, [11] introduced a seismic-based sleep monitoring system designed for on-bed sleep scenarios, showing impressive real-time monitoring capabilities for heart rate, respiration rate, and inter-beat interval, achieving good per-

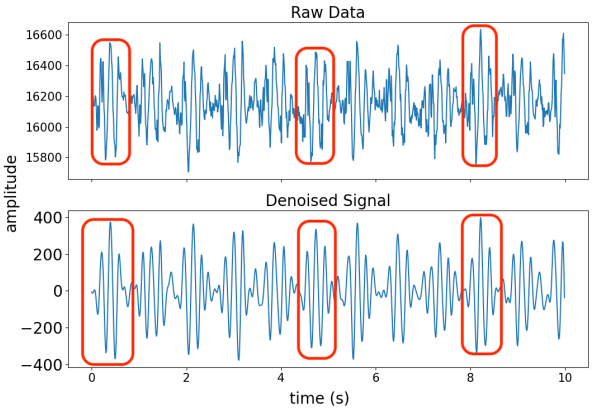

Fig. 4. Raw Data (Top) and Denoised Signal (bottom) for HR estimation.

formance. Similarly, [12]–[14] have developed analogous on-bed systems primarily to monitor HR and RR. It's noteworthy that the majority of these previously proposed systems have focused on bed scenarios and have relied on adaptive signal processing methods, which, while effective, may not be as robust as data-driven approaches such as deep learning.

[11] notably demonstrated the transferability of proposed approach from bed to seat scenarios, yielding favorable outcomes. However, it's worth noting that deep learning is not applied in this work, which have the potential to significantly enhance vital signs estimation results. Therefore, there is substantial potential for exploring deep learning methods to further enhance performance of seat scenario.

III. Methodology

The workflow of the SeatDot model is illustrated in Fig. 2. In this section, we introduce the training and inference methodologies of the SeatDot model. It's important to highlight that we trained two distinct models: one tailored for HR estimation and the other for RR estimation, each necessitating unique feature sets.

A. Extracting Heartbeat and Respiration signal

Before training the models for HR and RR estimation, it's essential to extract the desired signals for each. Typically, the desired frequency range for HR lies between 0.6 Hz and 3 Hz, while for RR, it is between 0.1 Hz and 0.6 Hz. However, in our study, the recorded seismic sensor signal for heartbeats does not resemble a pure sine wave. Instead, it presents more of a 'W' shape for a single heartbeat cycle, as highlighted by the red rectangles in the top panel of Fig. 4. This pattern leads to an approximate tripling of the corresponding frequency range. Consequently, we applied a Butterworth bandpass filter with a cutoff frequency ranging from 2 Hz to 10 Hz to the raw data, evident in the bottom panel of Fig. 4.

When considering the respiration signal, although the desired frequency range for the respiration rate (RR) is below 2 Hz, applying a low-pass filter with a 2 Hz cutoff directly to the raw data can cause distortions. This distortion arises due to the low signal-to-noise ratio, as illustrated in the last row

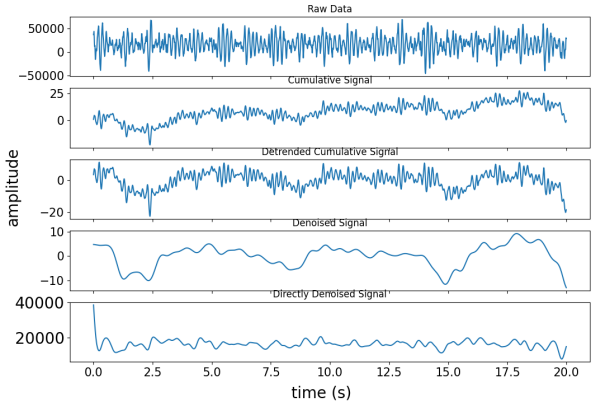

Fig. 5. Raw Data (first row), Directly applied low pass filter on raw data (last row) and Respiration envelope extraction process in the middle three rows.

of Fig. 5. As discussed in [11], the raw signal comprises the heartbeat signal, respiration signal, and random background noise. To extract the respiration signal pattern, the researchers integrate the raw velocity data, which is highlighted in the second row of Fig. 5. This integration is necessary because the respiration signals result from chest movements, which are better represented by the displacement signal rather than the velocity signal. However, the baseline of this cumulative signal tends to drift due to system errors, with the variance of such errors increasing over time [11]. To rectify this, the cumulative signal is detrended, as displayed in the third row of Fig. 5. After detrending, one can apply either a low-pass filter or wavelet denoising to achieve a cleaned respiration signal, as shown in the second-to-last row of Fig. 5.

B. Self-Supervised Contrastive Representation Learning

Fig. 2 provides an overview of the architecture used in SeatDot. We begin with denoising the raw data to extract both the heartbeat signal (Xh_i) and respiration signal (Xr_i). These signals are then transformed into two different views: (Xh_i) becomes (Xh_i^1, Xh_i^2), and (Xr_i) is converted to (Xr_i^1, Xr_i^2) through data augmentations. For ease of simplicity, we will refer to both (Xr_i) and (Xh_i) as (x_i) henceforth, given that the training methodology for both signals is the same. Next, (x_i^1) and (x_i^2) are processed through an encoder, which extracts representations and then maps them into new embeddings (z_i^1 and z_i^2). This mapping utilizes a projector that consists of two fully connected layers, housed within the self-supervised representation learning module (as illustrated by the green rectangle in Fig. 2). Finally, a contrastive loss function is applied on the pair of projected embeddings . This function aims to maximize the mutual information between (z_i^1) and (z_i^2) while minimizing the mutual information between different data samples ($x_i \& x_j$ with $i \neq j$). Contrastive learning is a well-known technique for learning effective representations by creating diverse views of the same data and simultaneously maximizing the shared information between these views. Re-

markably, this training approach can be accomplished entirely in a self-supervised manner, negating the need for labeled data.

1) Data augmentations: We apply two data augmentation techniques to transform the input data samples, helping to discover invariant features through contrastive learning. It's important to note that the features for both HR and RR estimation should exhibit periodic patterns. Therefore, our data augmentations must preserve the underlying periodicity of the input. To meet this criterion, we have several options: (a) Horizontal or Vertical Flipping: This method flips the input data either horizontally or vertically, depicted in the second and third rows of Fig. 6. (b) Scaling: Another approach is adjusting the amplitude of the input data, illustrated in the fourth row of Fig. 6. (c) Adding Gaussian Noise: We can introduce Gaussian noise with a mean of zero and a specific variance, as demonstrated in the last row of Fig. 6. These augmentation techniques are selected to ensure the preservation of the periodic nature of the input data, which is paramount for precise HR and RR estimations.

2) Contrastive Representation Learning: The architecture of the encoder is shown in Fig. 2. Our input data has dimensions of 1×1000, where "1" denotes a single sensor, and "1000" represents the length of the input signal, which corresponds to 100 Hz data over 10 seconds. The output representations generated by the encoder are of dimensions 256 with a temporal length of 31. The operation of the encoder network is introduced as follows: The input signal is first processed through the encoder, which consists of five temporal residual blocks, as detailed in [15] and [16]. Unlike regular convolution, these residual blocks utilize dilated convolution [17] to enhance their performance. Each temporal residual block, as shown in Fig. 3, comprises two dilated convolution layers. The first dilated convolution layer includes a stride of 2 to perform down-sampling. This entire process does not rely on labeled data. Instead, the encoder aims to learn invariant representations by processing different views of the same input through data augmentation. These learned representations will later be fine-tuned using a smaller set of labeled training data.

The encoded representations go through a projector, as illustrated in Fig. 2, which produces a projected output denoted as z_i. To enhance the similarity between two differently augmented views, namely $z^1 i$ and $z^2 i$ of the same input, we employ a contrastive loss function. This loss function operates on the projected embeddings z_i, indirectly refining the representations obtained from the encoder. These refined representations will be utilized in later phases for prediction. The mathematical expression for the contrastive loss is as follows:

$$\ell_i^c = -\log \frac{\exp\left(\text{sim}\left(z_i^1, z_i^2\right)/\tau\right)}{\sum_{j=1}^{2N} \mathbb{1}_{i \neq j} \exp\left(\text{sim}\left(z_i, z_j\right)/\tau\right)}, \tag{1}$$

where $\text{sim}(u, v) = u^T v/(\|u\| \cdot \|v\|)$, and z_i^1 and z_i^2 represent the outputs of the projector, which are obtained by processing different augmented views of the same input, as depicted in Fig. 2. Given a batch of N input examples, we obtain twice that number in augmented samples, resulting in $2N$ samples. For each sample, the two data augmentation views are considered as a positive pair. Conversely, for any given sample x_i and its counterpart x_j, where $i \neq j$, is treated as a negative pair. In this context, τ denotes a temperature parameter, and $sim(u, v)$ represents the cosine similarity between vectors u and v.

C. Supervised Representation Learning

Self-supervised contrastive representation learning aims to capture invariant representations of data in instances level by leveraging different views of the same input through contrastive learning. In contrast, supervised representation learning refines these representations to align more closely with the actual ground truth labels. As inspired by [18], the goal is to establish a strong connection between the feature representations and the labels, ensuring that the learned features are strongly associated with their corresponding labels. During the supervised learning phase, we utilize a small set of labeled data to guide these features toward a dedicated vital sign feature space. This is achieved by establishing the following relationship:

$$\ell_i^s = \|d(z_i, z_j) - |y_i - y_j|\,\|_2^2, \tag{2}$$

where y_i and y_j are labels of x_i and x_j, and $d(\cot)$ is the euclidean distance function. When the labels for a pair of samples have a smaller difference, their representations in the latent space become closer to each other. This means that the representation manages to preserve the consistent features unique to each instance while also maintaining a strong connection to the label space.

D. Training and Inference Stage

Our training process consists of two stages: pre-training and fine-tuning. During the pre-training stage, we use unlabeled data to pre-train the model. We achieve this by applying self-supervised contrastive learning, as previously described. In the subsequent fine-tuning stage, we introduce two fully connected layers on top of the pre-trained encoder network. This fine-tuning is carried out in a supervised manner for regular supervised learning. The projector used during training is also employed for additional supervised representation learning, as detailed in the previous section. This further adjusts the representations towards the label space. During the fine-tuning, only 5% of labeled training data is needed to achieve similar performance as fully supervised learning which typically requires 100% labeled data for both HR and RR estimation. The overall loss function for the fine-tuning stage is as follows:

$$\ell = \sum_{i=1}^{N} \|\hat{y_i} - y_i\|_2^2 + \lambda * \ell_i^s. \tag{3}$$

During the inference stage, we only need the encoder and the two fully connected layers, which are highlighted within the red rectangle in Fig. 2.

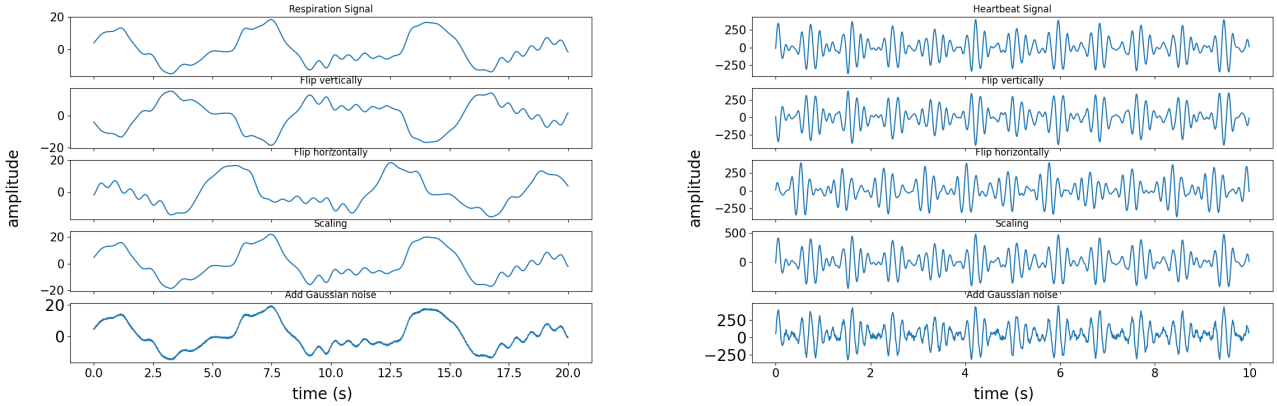

Fig. 6. Data augmentations for respiration signal (left) and heartbeat signal (right)

IV. Experiment Setup and Performance Evaluation

We conducted our experiments and evaluations on 32 subjects seated in a chair, with each recording session lasting between 3 to 5 minutes. Additionally, we collected data from from an additional 60 subjects while they were lying in bed[1] as training dataset. For data labeling, we utilized an FDA-approved oximeter, the CareTaker4, which offers continuous, real-time estimations of vital signs, including heart rate (HR) and respiration rate (RR). In our evaluations, we present metrics such as the Mean Absolute Error (MAE), the Standard Deviation of Absolute Error (STD), and the Mean Absolute Error Percentage (MAPE) for each conducted experiment.

A. Implementation Details

We use data collected from bed as training dataset and data collected from seat as testing dataset. For HR estimation, we used 10-second segments. For RR estimation, we employed 20-second segments. Importantly, we ensured that there was no overlap in subjects between the training and testing datasets. This strategy was crucial to demonstrate the effectiveness of our method in generalizing across different individuals and is able to transfer seamlessly from bed to seat. Our evaluation involved reporting performance metrics, including MAE, STD and MAPE for both our proposed methods and baseline methods. During training, we utilized a batch size of 128 and employed the Adam optimizer with a learning rate of 0.001, weight decay of 0.0003, $\beta_1 = 0.9$, and $\beta_2 = 0.99$. We found that setting $\lambda = 0.1$ produced good results during our experiments. Additionally, we applied a dropout rate of 0.1 in every residual block. For the HR model, the kernel size of convolution layers in residual blocks was set to 3, while for the RR model, it was set to 9.

B. Compare with Baseline methods

We compare our proposed approach against the following baselines. (a) Seismic-based System [11]: We contrasted our technique with a previously developed seismic-based system designed specifically for sleep monitoring in a bed setting. However, this system also demonstrated potential in transitioning from bed to seat. (b) Fully Supervised Learning: Our method was evaluated against Fully Supervised Learning. Here, supervised learning was applied to both the encoder and the fully connected layers. This approach requires 100% labeled training data collected from the bed. To evaluate the performance of our model, we train two fully connected layers on top of a self-supervised pre-trained encoder model. In table I, you can find the results of our approach compared to the baseline methods. In general, our proposed method surpasses all the baseline methods in terms of MAE, STD, and MAPE for either HR or RR estimation.

TABLE I
COMPARE WITH BASELINE METHODS

METHOD	VITAL	MAE	MAPE	STD
[11]	**HR(BPM)**	3.88	0.0497	8.82
	RR(BPM)	3.37	0.2427	2.83
100% FULLY SUPERVISED	**HR(BPM)**	3.33	0.0457	4.71
	RR(BPM)	2.28	0.1904	1.69
100% PROPOSED	**HR(BPM)**	**2.44**	**0.0333**	**4.30**
	RR(BPM)	**2.20**	**0.1826**	**1.57**

C. Compare with Fully Supervised Learning

In this analysis, we explore the efficacy of our proposed model in scenarios with limited training samples. Experiments were conducted using both small subsets (5% for HR and 15% for RR) and the entirety (100%) of the available training data. These results were then benchmarked against those obtained from fully supervised learning for HR and RR estimation. Remarkably, our model consistently surpassed the performance of supervised learning in both limited and full training data scenarios. Furthermore, it's worth emphasizing that our method, even when leveraging only 5% (for HR) or 15% (for RR) of the training data, yielded comparable results in terms of MAE and MAPE compared to fully supervised learning that used the entire training dataset. Such findings highlight the robustness of our approach, especially in contexts where access to extensive labeled training data is constrained, as presented in Table II.

[1]These clinical experiments received approval from the IRB under PROJECT00001838.

TABLE II
COMPARE WITH SUPERVISED LEARNING IN DIFFERENT CASES

METHOD	VITAL	MAE	MAPE	STD
5%/15% FULLY SUPERVISED	**HR(BPM)**	5.32	0.0685	8.06
	RR(BPM)	2.58	0.2095	1.90
5%/15% PROPOSED	**HR(BPM)**	3.25	0.0417	5.74
	RR(BPM)	2.22	0.1848	**1.53**
100% FULLY SUPERVISED	**HR(BPM)**	3.33	0.0457	4.71
	RR(BPM)	2.28	0.1904	1.69
100% PROPOSED	**HR(BPM)**	**2.44**	**0.0333**	**4.30**
	RR(BPM)	**2.20**	**0.1826**	1.57

D. Ablation Study

We conducted a detailed investigation to assess the impact of each component within our proposed methodology. or a thorough comparison, we designed several model variants, operating under the premise that only 5% of the labeled training data was available for HR estimation and 15% for RR estimation. We examined the following scenarios. (a) **Vanilla**: This model was trained entirely from scratch. We used 5% of the labeled training data for HR estimation and 15% for RR estimation, employing a fully supervised approach (b) **Pre-train**: For this variant, we pre-trained the model using all the available training data, without considering any label information during this phase. After this, we introduced a regressor on top the pre-trained encoder, using just 5% of the labeled data for HR and 15% for RR. (c) **All**: This model embodies our comprehensive approach. We began by pre-training the model using the entirety of the training data, omitting label information. This was followed by the introduction of a regressor on top of the pre-trained encoder, combine with supervised representation learning. The outcomes from this comparative study are showed in Table III. These results showcase that an integrative approach, combining all the components, yields the best performance.

TABLE III
ABLATION STUDY UNDER LIMITED TRAINING DATA

METHOD	VITAL	MAE	MAPE	STD
VANILLA	**HR(BPM)**	5.32	0.0685	8.06
	RR(BPM)	2.58	0.2095	1.90
PRE-TRAIN	**HR(BPM)**	3.49	0.0451	5.88
	RR(BPM)	2.36	0.1936	1.66
ALL	**HR(BPM)**	**3.25**	**0.0417**	**5.74**
	RR(BPM)	**2.22**	**0.1848**	**1.53**

V. CONCLUSION

In this paper, we introduce the SeatDot system, a novel approach that combines self-supervised contrastive representation learning and supervised representation learning to monitor vital signs using a seismic sensor. Our tests encompassed 32 subjects and achieved impressive results: an MAE of 2.44 for HR estimation and 2.20 for RR estimation. When compared with an FDA-approved device, these outcomes align with or even surpass state-of-the-art performance. Notably, SeatDot can achieve comparable results using merely 5% (for HR) and 15% (for RR) of the labeled training data in comparison to a fully supervised model relying on 100% labeled data. This is evident across metrics like MAE, STD, and MAPE. Furthermore, no overlap of subjects between training and testing datasets underscores the ability of proposed method in generalizing across different users. Overall, SeatDot showcases immense promise for integration into smart office settings, aiming to monitor well-being of humans during work hours.

REFERENCES

[1] I. Sadek, E. Seet, J. Biswas, B. Abdulrazak, and M. Mokhtari, "Non-intrusive vital signs monitoring for sleep apnea patients: a preliminary study," *IEEE Access*, vol. 6, pp. 2506–2514, 2017.

[2] Y. Zhu, J. Maniyeri, V. F. S. Fook, and H. Zhang, "Estimating respiratory rate from fbg optical sensors by using signal quality measurement," in *2015 37th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. IEEE, 2015, pp. 853–856.

[3] M. Krej, Ł. Dziuda, and F. W. Skibniewski, "A method of detecting heartbeat locations in the ballistocardiographic signal from the fiber-optic vital signs sensor," *IEEE journal of biomedical and health informatics*, vol. 19, no. 4, pp. 1443–1450, 2015.

[4] T. Rahman, A. T. Adams, R. V. Ravichandran, M. Zhang, S. N. Patel, J. A. Kientz, and T. Choudhury, "Dopplesleep: A contactless unobtrusive sleep sensing system using short-range doppler radar," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2015, pp. 39–50.

[5] J. Liu, Y. Chen, Y. Wang, X. Chen, J. Cheng, and J. Yang, "Monitoring vital signs and postures during sleep using wifi signals," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 2071–2084, 2018.

[6] G. Y. Lui, D. Loughnane, C. Polley, T. Jayarathna, and P. P. Breen, "The apple watch for monitoring mental health–related physiological symptoms: Literature review," *JMIR Mental Health*, vol. 9, no. 9, p. e37354, 2022.

[7] C. Paradiso, F. Colino, and S. Liu, "The validity and reliability of the mi band wearable device for measuring steps and heart rate," *International Journal of Exercise Science*, vol. 13, no. 4, p. 689, 2020.

[8] R. N. Aurora, S. P. Patil, and N. M. Punjabi, "Portable sleep monitoring for diagnosing sleep apnea in hospitalized patients with heart failure," *Chest*, vol. 154, no. 1, pp. 91–98, 2018.

[9] N. Molinaro, E. Schena, S. Silvestri, and C. Massaroni, "Multi-roi spectral approach for the continuous remote cardio-respiratory monitoring from mobile device built-in cameras," *Sensors*, vol. 22, no. 7, p. 2539, 2022.

[10] C. Yang, M. Hu, G. Zhai, and X.-P. Zhang, "Graph-based denoising for respiration and heart rate estimation during sleep in thermal video," *IEEE Internet of Things Journal*, vol. 9, no. 17, pp. 15 697–15 713, 2022.

[11] Y. Song, B. Li, D. Luo, Z. Xie, B. G. Phillips, Y. Ke, and W. Song, "Engagement-free and contactless bed occupancy and vital signs monitoring," *IEEE Internet of Things Journal*, 2023.

[12] Z. Jia, A. Bonde, S. Li, C. Xu, J. Wang, Y. Zhang, R. E. Howard, and P. Zhang, "Monitoring a person's heart rate and respiratory rate on a shared bed using geophones," in *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems*, 2017, pp. 1–14.

[13] Z. Jia, M. Alaziz, X. Chi, R. E. Howard, Y. Zhang, P. Zhang, W. Trappe, A. Sivasubramaniam, and N. An, "Hb-phone: a bed-mounted geophone-based heartbeat monitoring system," in *2016 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. IEEE, 2016, pp. 1–12.

[14] J. Clemente, M. Valero, F. Li, C. Wang, and W. Song, "Helena: Real-time contact-free monitoring of sleep activities and events around the bed," in *2020 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 2020, pp. 1–10.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[16] C. Lea, M. D. Flynn, R. Vidal, A. Reiter, and G. D. Hager, "Temporal convolutional networks for action segmentation and detection," in *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 156–165.

[17] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *arXiv preprint arXiv:1511.07122*, 2015.

[18] Z. Taghiyarrenani, S. Nowaczyk, S. Pashami, and M.-R. Bouguelia, "Multi-domain adaptation for regression under conditional distribution shift," *Expert Systems with Applications*, vol. 224, p. 119907, 2023.