SeismoDot: Self-Supervised Representation Learning for Occupancy Detection

Bed occupancy detection is a fundamental task in automated sleep monitoring systems, serving as the initial step in identifying an individual's presence in bed before proceeding with subsequent tasks. However, many existing approaches for bed occupancy detection are constrained to controlled environments characterized by uniform room setups, backgrounds, and beds. Furthermore, many of these methods rely on threshold-based approaches, which may not be optimal and robustness compare with data-driven methods such as deep learning. Especially, their performance degrades rapidly when tested in new, unseen environments due to distribution shift in background noise. Consequently, the challenge of adapting bed occupancy detection algorithms to diverse and novel settings remains largely unexplored. Also, collecting extensive data from various environments to facilitate robust model generalization is exceptionally challenging, particularly for sensory data, where public datasets are scarce, unlike in the field of computer vision. Moreover, many existing sensor data-related tasks only leverage either the temporal or frequency domain features, neglecting the potential of leveraging both features altogether. However, numerous tasks, such as human activity recognition (HAR), irregular heartbeat classification (AF detection), and occupancy detection, etc. involve both temporal and frequency features. In this paper, we present a novel bed occupancy detection algorithm that leverages self-supervised learning which includes multi-class classification module and spectrum-temporal feature fusion module using data collected from seismic sensors. Our proposed algorithm force model learn a robust representation capable of generalizing effectively across various environments, even with limited training data. This approach has been comprehensive evaluated across diverse environments, achieving an overall accuracy rate of 98.56% and F1 score of 97.57% in the task of distinguishing between on-bed and off-bed statuses in various different environments. Our experimental results demonstrate the adaptability and robust performance of the algorithm across different environments. Even when trained with just 20% of the available training data, the performance is comparable to models trained on the full dataset.

CCS Concepts: • **Human-centered computing → Ubiquitous and pervasive computing**.

Additional Key Words and Phrases: Bed Occupancy, Self-Supervised Learning, Spectrum-temporal feature fusion

ACM Reference Format:
. 2023. SeismoDot: Self-Supervised Representation Learning for Occupancy Detection. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 1 (October 2023), 11 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Contactless sleep monitoring is showing great promise as an alternative to wearable sleep devices. Recently, numerous research in this area have emerged, with a particular focus on contactless vital signs estimation which is crucial for studying various sleep-related disorders. However, bed occupancy detection which is the fundamental task among many sleep studies. Bed occupancy detection is the first step for all sleep monitoring system to perform subsequent tasks. while nontrivial, has not received the attention it deserves.

Many existing works utilize bed sensors such as mat pressure sensors [10, 14] and accelerometers [1, 9] to perform bed occupancy. The former may not very easy to install and the latter is easy to install but may not as sensitive as seismic sensors when it comes to capture physiological signal from human in contactless

Author's address:

2474-9567/2023/10-ART $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

manner. Compare to most accelerometers (micro-g sensitivity), the seismic sensor has a significantly higher sensitivity (nano-g sensitivity) level and can capture higher fidelity signals. Other existing works such as [21] using cameras may cause privacy concerns, lead to uncomfortable feelings especially during sleep period. Furthermore, many of these studies on bed occupancy predominantly concentrate on single environments, characterized by same room setups and uniform bed types. Consequently, when subjected to changes in environment, the performance of bed occupancy algorithms tends to decline due to shifts in distribution of background noise. This issue is particularly pronounced in off-bed detection due to background noise is off bed period, which significantly affecting the accuracy of off-bed period identification. Many existing methods today need a re-calibration process, which involves tweaking a predetermined threshold, every time they're used in a new environment. But making these methods adapt by adjusting the threshold requires having labeled data in that new environment, which isn't convenience in many real-world situations from a user perspective. In addition, The environment distribution might be changed with time which could require further adjustment of threshold. Moreover, finding the best threshold of adaptive statistical signal processing algorithm is challenging in various complex environment settings. This highlights the importance of using a data-driven approach to learn the best 'boundary' for distinguishing between off-bed and on-bed signals.

Unlike image data, which benefits from a large amount of publicly available datasets for training generalized deep learning models, sensor data often lacks access to large, open training datasets. Collecting substantial real-world data from diverse environments and multiple patients is a challenging and time-consuming endeavor, often spanning months or even years. Additionally, many sensor-related tasks involve features from both the spectral and temporal domains, such as HAR, AF detection, or occupancy detection. While some existing approaches exclusively utilize temporal domain features, and others focus solely on spectral domain features, the potential advantages of leveraging both domains remain under-explored. In this paper, we propose a self-supervised representation learning method that combines self-supervised learning with spectrum-time feature fusion, significantly reducing the requirement for vast amounts of data to train a robust model. Our proposed system, SeismoDot, demonstrates the ability to accurately detect bed occupancy with just 20% of labeled training data across various environments.

We have successfully deployed SeismoDot for real-time bed occupancy detection across various environmental settings. The system's design and key components are visually represented in figure 1. SeismoDot comprises a Raspberry Pi serving as the onboard computer, an Analog Digital Converter (ADC) board efficiently digitizes data at a sampling rate of 100 Hz, and a vertical seismic sensor (seismometer) designed to detect vibrations. This seismic sensor incorporates a magnetic element surrounded by wire coils, and as the magnetic element moves within these coils, it generates an electrical signal. The seismic sensor exhibits a remarkable sensitivity to micro-vibrations stemming from human cardiac activity, surpassing the capabilities of most MEMS accelerometers [5], enabling SeismoDot to capture real-time cardiac activity without requiring physical contact. Furthermore, SeismoDot has magnetic mounted on it which makes it easy to install as shown in the left of figure 1.

Existing bed occupancy detection methods, such as [1, 10, 14], require re-initialization for each new environment, which may not be practical in real-world applications compared to methods that do not require re-calibration. The approach presented in [2] focuses solely on bed entry and exit events, which can lead to cascading errors if either of these detections fails. Other methods like [9] and [13] require back correction, with [9] involving multiple stages of classification and feature extraction, potentially resulting in information loss. While both [7] and [13] continuously detect bed occupancy, they rely on statistical signal processing techniques, which may not be as robust as data-driven methods like deep learning. In addition, [7] may experience a significant performance drop when tested in more complex environments. In contrast, SeismoDot distinguishes itself in the following ways: (1) It performs bed occupancy detection continuously using a deep learning model without the need for back corrections. (2) It demonstrate robustness to changes in the environment and requires no re-initialization when deployed in new settings. (3) It does not demand extensive training data to adapt to various environments.

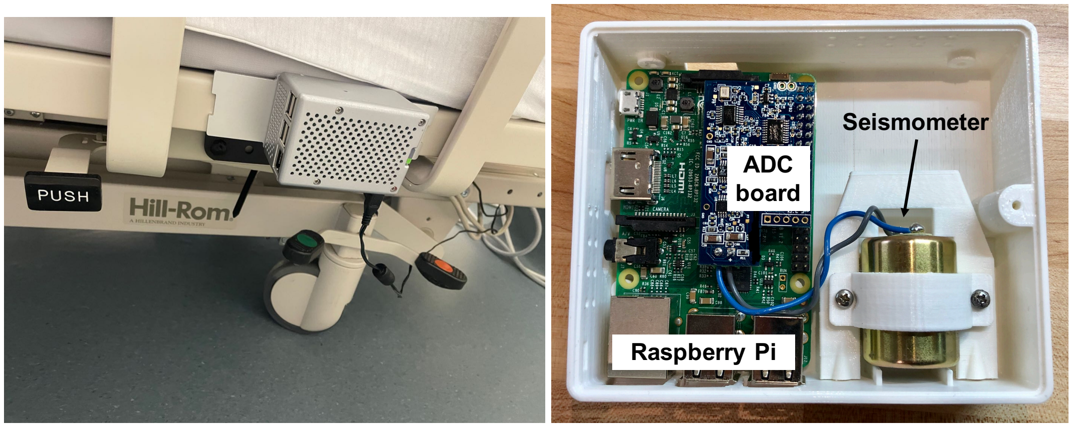

Fig. 1. Installation of SeismoDot on the side of bed frame(left), Design of SeismoDot(right).

While threshold-based or statistical signal processing methods do not require training data, they often require exposure to significant data volumes to fine-tune thresholds for optimal performance.

To achieve accurate bed occupancy detection, several technical challenges has been posted as outlined following: (1) Many existing bed occupancy detection methods primarily target identifying the 'moment of bed entry' and 'moment of bed exit.' However, relying solely on these events can introduce significant errors if false detections occur even once. Simple amplitude-based checks to detect bed occupancy are prone to failure due to variations in baseline values across different environments. While bed entry and exit typically produce higher signal energy, excessive movements on the bed can trigger false detections and potentially lead to cascading errors. (2) Developing a robust model that maintains high performance across different environments which requires training the model to learn effective representations that can adapt to varying conditions. (3) From a user's perspective, it's essential that the bed occupancy detection device is easy to install and can be deployed without requiring repeated initialization steps when placed in new environments. User-friendliness and seamless deployment are key considerations in the design of such systems.

We have developed solutions to tackle these challenges in bed occupancy detection. SeismoDot employs a deep learning model built upon self-supervised representation learning for tackling bed occupancy detection in limited training data situation. Our goal is to learn representations that can generalize to more complicated new environments with limited training samples. The self-supervised learning scheme comprises two key modules: multi-class classification and spectral-temporal feature fusion. The entire self-supervised learning task is treated as an auxiliary objective and does not require pre-training. Instead, it is jointly optimized with the primary bed occupancy detection task. This approach is different from the prevalent strategy found in current literature, which typically involves pre-training on the entirety of unlabeled data and subsequent fine-tuning, requiring the freezing of the pre-trained feature extractor and training only on the final layers for classification. We opted for this alternative approach there is no guarantee that the learned representations during the pre-training stage is highly related to downstream tasks. The multi-class classification module is to efficiently recognize various transformations applied to both temporal and spectral features, enabling the learning of invariant features across different transformations [11]. For feature fusion module, which explicitly leveraging adversarial training, to create fine-grained, rich representations. While we assume an overlap between temporal and spectrum features, our feature fusion strategy seeks diversity rather than homogeneity. To achieve this, we maximize the diversity of temporal and spectrum features by incorporating a discriminator. Simultaneously, we maximize the similarity between the two types of features through contrastive learning to ensure there is meaningful overlap (invariant

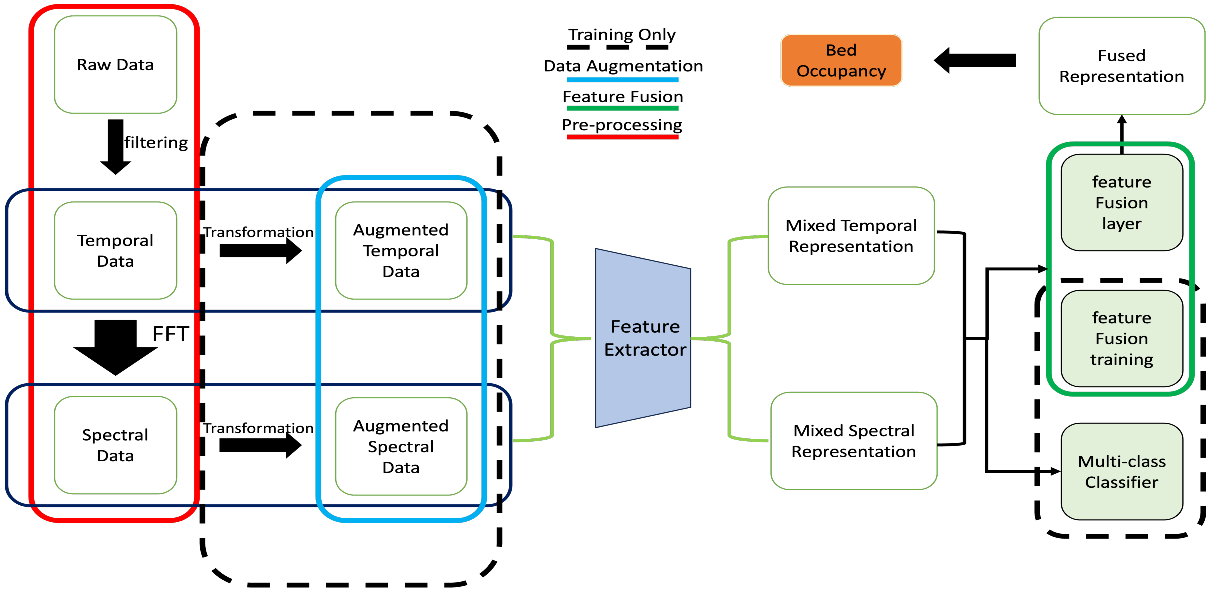

Fig. 2. Workflow of SeismoDot.

features across both time domain and frequency domain) between temporal and spectrum representations. Lastly, SeismoDot is designed to be compact and easy to install, enhancing its practicality and usability in various settings.

Figure 2 provides an overview of the operational flow in SeismoDot. It begins with the pre-processing of data in time-doamin using a band-pass filter. Subsequently, data augmentation is applied to input for both the time-domain and spectral domain of data, combined with multi-class classification, to generate efficient representations. Following this, the representations is being further refinement via the spectrum-temporal feature fusion module. In the meanwhile, the model is under supervised training using a small portion of labeled training data.

We outline the **key contributions** of this paper as follows:

(1) We introduce a self-supervised learning module includes multi-class classification in conjunction with a spectrum-temporal feature fusion module, achieving performance utilizing only 20% of labeled training data which is comparable to fully supervised learning on full dataset .
(2) We propose a spectrum-temporal feature fusion module that leverages information from both the frequency domain and time domain, demonstrating its effectiveness through experiments.
(3) Our comprehensive evaluation, conducted across various users and diverse environments, yields an average accuracy of 98.56% and an F1 score of 97.57%.
(4) Our proposed system could be deployed in real-world environments, enabling continuous real-time bed occupancy detection.

2 RELATED WORKS

2.1 Bed Occupancy Detection

[1, 10, 14] calculated a baseline threshold when no one is on bed as the prior knowledge. Then, the calculated threshold is used to compare with incoming signals to determine the bed occupancy status. This type of methods are easy and intuitive. However, there is no statistical result for this method on those references which makes it hard to evaluate the performance. In addition, it needs to be re-initialized each time with a different environment.

[2] proposed a feature fusion method by combining Spectral Entropy, Kurtosis and Teager Energy Operator (TEO) to recognize bed entry and bed exit. They achieved high accuracy on their dataset. However, they only detect moments of 'bed entry' and 'bed exit' instead of continuously detecting 'on bed' or 'off bed' status, that potentially could bring continuous errors if one moment of 'bed entry' and 'bed exit' has been falsely detected.

[7] using an autocorrelation function to detect if the signal is periodic to differentiate 'on bed' and 'off bed' signals. This method could achieve very high accuracy. However, when another periodic signal other than a heartbeat exists it could lead to false detection.

In the study by [9], bed occupancy detection is performed using feature extraction and Long Short-Term Memory (LSTM) techniques with accelerometer data. Their approach involves the design of three distinct classifiers: one for detecting stage changes, another for identifying intervals between stage changes, and a third for making corrections based on the outputs of the first two classifiers, ultimately achieving good performance. However, it's important to note that their testing was confined to the same environmental conditions, encompassing only seven patients. Furthermore, their handcrafted feature extraction process yielded 21 dimensions, with a majority being spectrum magnitudes which may lose of information.

[13] proposed a signal processing method which combines zero crossing rate and kurtosis to continuously detect on bed and off bed and achieves a very high accuracy. However, it requires back correction as movement could potentially affect the off bed detection.

2.2 Self-supervised Representation Learning for Time-Series

Contrastive learning, as described by [8], aims to learn invariant representations by maximizing the mutual information between past representations and future representations. This approach involves comparing similar and dissimilar pairs of data points and encouraging the network to produce similar representations for the former and dissimilar representations for the latter. By maximizing the mutual information between these pairs, contrastive learning enables the network to learn robust and generalizable features that can be used for a range of downstream tasks.

In a recent work by Zhang et al. [20], they introduced a novel model named 'Crossformer,' built upon the transformer architecture. What sets this model apart is its ability to capture not only temporal dynamics but also the interplay between different modalities simultaneously. Their experiments showcased its impressive performance, and while the primary aim of their paper was to enhance forecasting results, it also contributes to representation learning—a fundamental aspect, as many existing studies leverage forecasting for representation learning.

Eldele et al. [3] proposed a contrastive learning framework focused on learning representations. They incorporate both temporal and contextual contrast aspects into their approach. In addition, they propose a cross view module further enhance representation learning. [11] introduced a multi-task self-supervised learning paradigm. This paradigm transforms raw data into various views and conducts binary classification on each transformation to learn robust representations. Sarkar et al. [12] build upon a similar idea and apply it to different tasks. Yue et al. [17] delve into self-representation learning for time-series data. They achieve this by generating diverse data views using techniques like masking and random cropping of time steps, facilitated by a hierarchical temporal contrast mechanism. Kiyasseh et al. [6] propose a contrastive learning framework that aims to learn invariant representations within individual patients. While promising for personalized health monitoring, its ability to generalize across different individuals may be limited. Tonekaboni et al. [15] tackle the challenge of estimating stationary temporal window sizes using the Augmented Dickey-Fuller (ADF) statistical test. They consider issues with naive negative sampling, which may include false negatives and adversely affect embedding learning, addressing this through Positive Unlabeled learning. Fan et al. [4] explore the relationships between different samples in the time domain and within the same sample but across different time segments. Lastly, Zhang et al. [18] propose a modified contrastive learning method known as Skip-Step Contrastive Predictive Coding, which includes a skip step to create a temporal gap between future and past representations.

These diverse approaches collectively contribute to advancing self-supervised learning and representation learning in various domains. However, The literature discussed thus far predominantly focuses on utilizing temporal information, with an emphasis on forecasting tasks where spectral information might not hold as much relevance as temporal features. Nonetheless, it is worth highlighting that spectral information remains crucial in tasks like occupancy detection.

On the contrary, [19] took into account both temporal and spectral domains in their approach. They introduced a contrastive learning method aimed at improving data representations by increasing the similarity between data from the same sample in both the time and frequency domains. They also projected time and frequency representations into a shared latent space to ensure alignment. However, the specific method they used to handle the frequency domain augmentation remains further discussion. Additionally, their focus was on aligning the time and frequency domains rather than fusing them.

2.3 Temporal-Spectral feature fusion

[16] suggested the utilization of dropout on input data as an instance-level augmentation technique. Their experiments indicated that dropout augmentation outperformed various other augmentation techniques. Furthermore, they introduced a bilinear temporal-spectral fusion module aimed at iteratively refining representations. They employed contrastive learning to maximize the alignment between temporal and spectral features. However, our approach differs from theirs, as we focus on promoting diversity between temporal and spectral features rather than making them homogeneous.

3 PRELIMINARY

In this section, we delve into the motivations behind our research on bed occupancy detection. We explain our preference for deep learning models over the 'feature extraction + machine learning' and statistical approaches. We begin by presenting the outcomes of our experiments with statistical signal processing, 'feature extraction + machine learning' strategies. Subsequently, we expound on our rationale for selecting deep learning as the preferred approach.

3.1 Statistical Signal Processing Approach

Initially, we explored strategies akin to those found in [7] and [13]. While [7] proved effective when the on-bed signal exhibited clear heartbeat patterns with a high Signal to Noise Ratio (SNR), it struggled when faced with noisy off-bed signals containing periodic noise. The algorithm in [7] directly identified periodic patterns using the Auto-correlation Function (ACF) on raw data, making it reliant on the periodicity of on-bed signals for detection. An example of a false off-bed detection is presented in Figure 3. Even though the signal is noisy, the periodic pattern is apparent, but the algorithm proposed in [7] failed to identify such signals. Moreover, external events near the bed during off-bed periods, as shown in Figure 3, caused false on-bed detections in both [7] and [13]. In the case of the algorithm introduced in [13], it heavily relies on the zero-crossing rate, inspired by speech detection in audio signals. Each time speech emerges after a period of silence, the zero-crossing rate drops due to the reduced fluctuations in heartbeat signals compared to background noise. However, it yields false detections in scenarios where off-bed signals exhibit constant low-frequency noise or even when the on-bed signal features clear heartbeat patterns but with a higher zero-crossing rate, as illustrated in Figure 3.

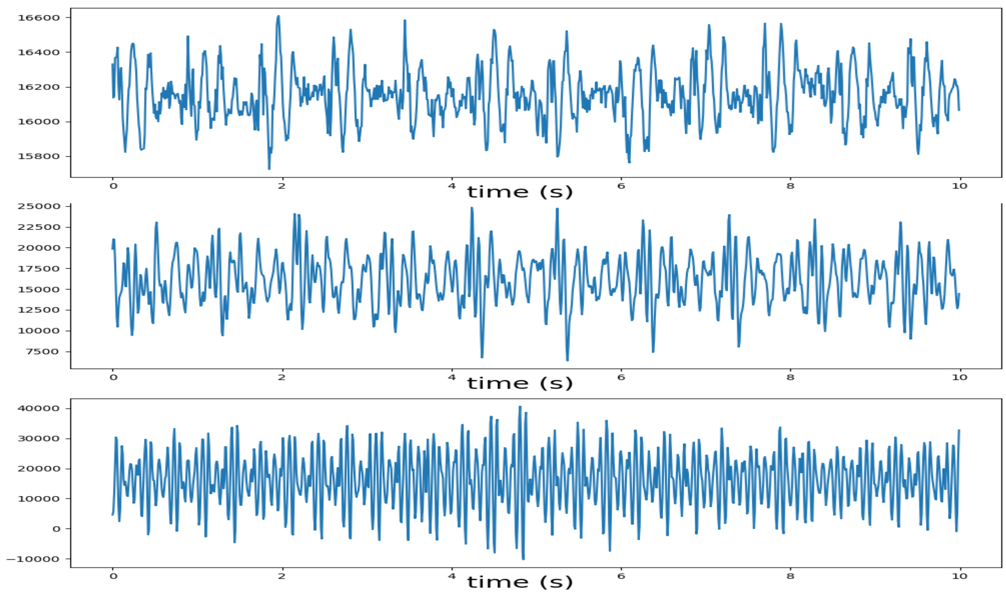

Fig. 3. Example of signal for false off bed (top), signal for false on bed (bottom) and signal for false off bed (middle) due to high zero crossing rate.

3.2 Traditional Machine Learning Approach

We conducted observations on the collected data, and in conjunction with the results presented earlier, we extracted 8 features and use random forest as classifier model.

3.2.1 Feature Extraction. We refer to the raw data as x and the band-pass filtered data as $\tilde{x} = filter(x)$. Additionally, we employed the envelope extraction method introduced in [13] and applied the Auto-correlation Function (ACF) on it, denoted as $\hat{x}_{ACF}$. Specifically, we extracted eight features as follows: (1) Primary frequency of raw data: This feature is computed as $argmax(FFT(x))$. Typically, the primary frequency for on-bed signals falls within the range of 1 Hz to 10 Hz, whereas off-bed signals tend to have frequencies exceeding 10 Hz. (2) Signal Fluctuation: Calculated as $max(x) - min(x)$. this feature measures the difference between the maximum and minimum values of the raw data. On-bed signals often exhibit larger difference on max and min amplitudes change compared to off-bed signals. (3) Dominant frequency after band pass filter denoised signal: Denoted as $argmax(FFT(\tilde{x}))$. this feature identifies the dominant frequency in the band-pass filtered denoised signal. Even after applying the band-pass filter, on-bed signals maintain a distinct dominant frequency compared to off-bed signals. In addition, the dominant frequency of filtered on-bed signal typically will stay the same compare to pre-filtered signal. However, for off-bed signal the dominant frequency is high likely to change after filtering. (4) Zero crossing rate of $\tilde{x}$: Calculated as $zcr(\tilde{x})$, where $zcr(\cdot)$ is zero crossing rate operation. this feature measures the rate of zero crossings in the band-pass filtered data. As previously mentioned, zero crossing rates tend to be higher in off-bed signals compared to on-bed signals. (5) Variance of mean from segmented ACF: First, we divided $\hat{x}_{ACF}$ into non-overlapping segments of equal length, denoted as $\{\hat{x}^i \in \hat{x}_{ACF} \mid 0 < i < N\}$. We then calculated the variance of the segmented ACF : $var[\{mean(\hat{x}^i)\}_{i=1}^{N}]$, where $var[\cdot]$ is variance operation. This feature typically exhibits a lower value for on-bed signals due to the periodic nature of these signals. The mean of periodic signals tends to be more stable compared to non-periodic signals. (6) Dominant Frequency of $\hat{x}_{ACF}$: Calculated as $argmax(FFT(\hat{x}_{ACF}))$. this feature identifies the dominant frequency of $\hat{x}_{ACF}$. $\hat{x}_{ACF}$ typically exhibits clearer

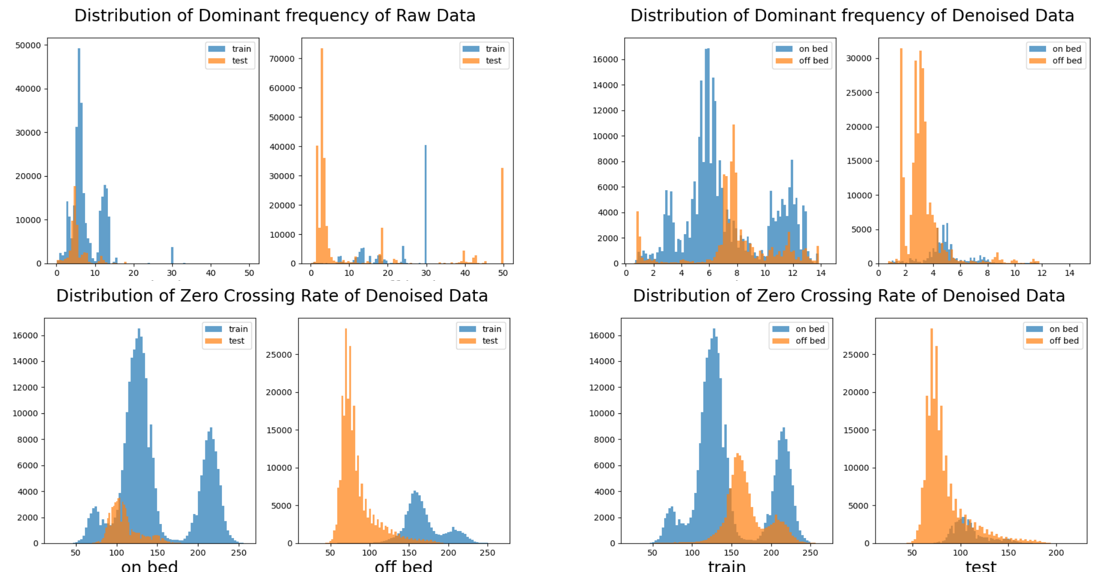

Fig. 4. Distribution of zero crossing rate and dominant frequency of raw data of both on bed and off bed signal on training and testing dataset

periodicity for on-bed signals with low SNR, making the dominant frequency a complementary information. (7) Max amplitude of spectrum calculated from $\hat{x}_{ACF}$: Measured as $max(FFT(\hat{x}_{ACF}))$. this feature characterizes the distribution of the spectrum of $\hat{x}_{ACF}$. On-bed signals tend to have a concentration of spectral amplitudes within a narrow range of frequency bins, resulting in a higher amplitude for the dominant frequency. Conversely, off-bed signals exhibit random spectral distributions and do not concentrate on a specific narrow bandwidth, leading to a lower amplitude for the corresponding dominant frequency. (8) Amplitude of max peak on $\hat{x}_{ACF}$: Represented as $max(\hat{x}_{ACF})$, this feature quantifies the periodic pattern within the given signal. The amplitude of the maximum peak in $\hat{x}_{ACF}$ is higher for periodic signals, indicating the similarity between the delayed signal and the original signal.

3.2.2 Feature Distribution Visualization. A significant challenge in bed occupancy detection arises from the distribution shift between the source environment (training data) and a new environment (testing data), particularly for off-bed data. This distribution shift can result in a drop in performance when a model trained on the source environment data is validated on data from a new environment. To illustrate this issue, we present examples in Figure 4, which shows the distribution shift between the training data (source environment) and the testing data (new environment) on the left side of figure 4.Furthermore, the distribution of both on-bed and off-bed data, in both training and testing, exhibits an overlap in their distribution, as depicted on the right side of figure 4. This overlap in distribution poses a challenge for the model in accurately distinguishing between on-bed and off-bed data, contributing to the complexity of the bed occupancy detection task in diverse environments.

3.3 Motivations

As detailed in Section 3.2.2, the features extracted from the data are, by themselves, insufficient to effectively distinguish between on-bed and off-bed instances. Deep learning, well known for its capacity to extract high-level representations within an embedding space, emerges as a vital component in this process. By applying deep learning, we can obtain domain-invariant features, enhancing the capacity to perform bed occupancy detection

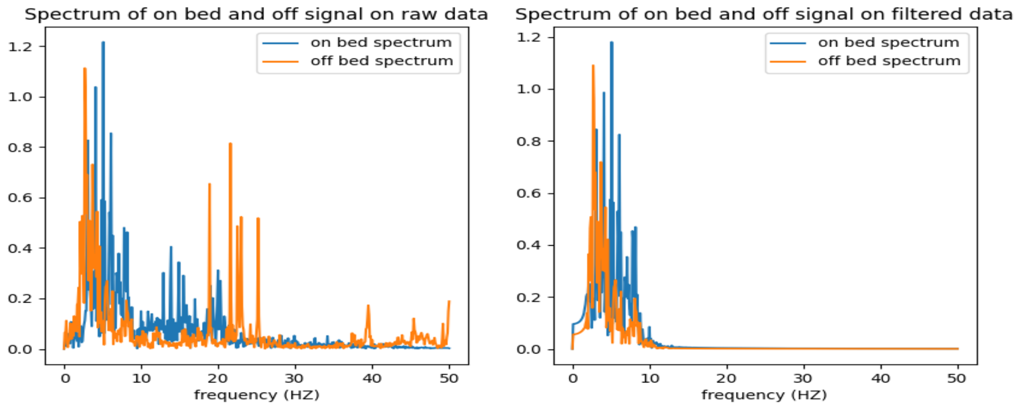

Fig. 5. Spectrum comparison between on bed and off bed data before and after filtering.

accurately and with greater generalizability. This proves to be particularly advantageous when compared to conventional methods, which rely on feature extraction combined with machine learning techniques. Consequently, the integration of deep learning for representation learning emerges as a key component of bed occupancy detection. Furthermore, as shown in section 3.2.1 and 3.2.2, it is imperative to consider both the time and frequency features in bed occupancy detection. On-bed data frequently exhibits periodic patterns, while on-bed and off-bed signals typically occupy different frequency ranges. This underscores the necessity of harnessing the capabilities of both the time and frequency domains to perform accurate bed occupancy detection. Lastly, collecting data from various individuals and diverse environments can be a time-consuming and challenging task. Hence, the adoption of self-supervised learning to extract robust representations is crucial for achieving high performance in bed occupancy detection, especially when dealing with limited training data.

4 METHODOLOGY

In this section, we first introduce pre-processing on raw data and data augmentation on both time domain data and spectral data. Then, we describe the self-supervised learning module includes both multi-class classification module on concatenation of temporal and spectral features and spectral-temporal feature fusion module.

4.1 Data Pre-processing

We implemented a band-pass filter with a cutoff frequency set between 2Hz and 10Hz on the time domain data. This choice is motivated by the fact that on-bed signals tend to predominantly occupy this frequency range, thus enhancing the visibility of on-bed features. Simultaneously, this filtering process has the potential to mitigate the influence of low-frequency noise present in the background noise and high frequency noise contains in on-bed signal. It's important to address the issue of low-frequency noise because it can contribute to false detections. This risk arises from the fact that on-bed signals also encompass respiration components, which can share the same frequency range as low-frequency noise. In addition, applying the band-pass filter within the 2Hz to 10Hz range does not significantly affect the background noise on time domain as it is randomly distributed across different frequency bins.

The spectral analysis of on-bed and off-bed signals reveals distinct patterns in the frequency components. To capture a more comprehensive representation, we conduct spectrum analysis on the raw data rather than the filtered data. This choice is motivated by the fact that raw data preserves richer patterns in the high-frequency part, which contains valuable information for distinguishing between on-bed and off-bed instances, as depicted on the left side of figure 5. In contrast, using spectrum analysis on filtered data can pose a challenge for discrimination because the frequency components tend to significantly overlap for both on-bed and off-bed signals, as illustrated on the right side of figure 5. We calculate the spectrum from 10-second raw data segments sampled at 100 Hz by employing Fast Fourier Transform (FFT). For our cases, we utilize only the first half of the spectrum, representing the positive frequencies, as our input.

4.2 Self-Supervised Learning

As Spectral-temporal feature fusion do not require label information, we consider it as self-supervised learning as well in this paper. In this subsection, we will discuss both multi-class classification and spectral-temporal feature fusion.

4.2.1 Multi-class Classification. We maximize mutual information between the future representation and the past representation for time domain features. We use data augmentation technique on spectrum first and then applying contrastive learning on it as well.

4.2.2 Spectral-temporal feature fusion. We try to force spectral and temporal features to be diverse in order to obtain rich information. In the meanwhile, we want to make sure there are some overlap between temporal and spectral embedding space but not too much. Therefore, the overall proposed module is to perform adversarial training.

5 EXPERIMENT

5.1 Dataset

We first collected data in a controlled hospital environment room. We collected 100 people's data in a hospital bed and family bed. We also collected 32 people's data when they are sit on a seat in that hospital environment. Furthermore, we collected 25 people's data with sensors on different location of the hospital bed (on the bottom and on the side).

REFERENCES

[1] Andreas Braun, Martin Majewski, Reiner Wichert, and Arjan Kuijper. 2016. Investigating low-cost wireless occupancy sensors for beds. In *Distributed, Ambient and Pervasive Interactions: 4th International Conference, DAPI 2016, Held as Part of HCI International 2016, Toronto, ON, Canada, July 17-22, 2016, Proceedings 4*. Springer, 26–34.

[2] Jose Clemente, Maria Valero, Fangyu Li, Chengliang Wang, and WenZhan Song. 2020. 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, 1–10.

[3] Emadeldeen Eldele, Mohamed Ragab, Zhenghua Chen, Min Wu, Chee-Keong Kwoh, Xiaoli Li, and Cuntai Guan. 2023. Self-supervised contrastive representation learning for semi-supervised time-series classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2023).

[4] Haoyi Fan, Fengbin Zhang, and Yue Gao. 2020. Self-supervised time series representation learning by inter-intra relational reasoning. *arXiv preprint arXiv:2011.13548* (2020).

[5] Zhenhua Jia, Amelie Bonde, Sugang Li, Chenren Xu, Jingxian Wang, Yanyong Zhang, Richard E Howard, and Pei Zhang. 2017. 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*. 1–14.

[6] Dani Kiyasseh, Tingting Zhu, and David A Clifton. 2021. Clocs: Contrastive learning of cardiac signals across space, time, and patients. In *International Conference on Machine Learning*. PMLR, 5606–5615.

[7] Fangyu Li, Maria Valero, Jose Clemente, Zion Tse, and Wenzhan Song. 2020. Smart sleep monitoring system via passively sensing human vibration signals. *IEEE Sensors Journal* 21, 13 (2020), 14466–14473.

[8] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. 2018. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748* (2018).

[9] Madhurananda Pahar, Igor Miranda, Andreas Diacon, and Thomas Niesler. 2023. Accelerometer-based bed occupancy detection for automatic, non-invasive long-term cough monitoring. *IEEE Access* 11 (2023), 30739–30752.

[10] Melanie Pouliot, Vilas Joshi, Rafik Goubran, and Frank Knoefel. 2012. Bed occupancy monitoring: Data processing and clinician user interface design. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 5810–5814.

[11] Aaqib Saeed, Tanir Ozcelebi, and Johan Lukkien. 2019. Multi-task self-supervised learning for human activity detection. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 2 (2019), 1–30.

[12] Pritam Sarkar and Ali Etemad. 2020. Self-supervised ECG representation learning for emotion recognition. *IEEE Transactions on Affective Computing* 13, 3 (2020), 1541–1554.

[13] Yingjian Song, Bingnan Li, Dan Luo, Zaipeng Xie, Bradley G Phillips, Yuan Ke, and Wenzhan Song. 2023. Engagement-Free and Contactless Bed Occupancy and Vital Signs Monitoring. *IEEE Internet of Things Journal* (2023).

[14] Matthew Taylor, Theresa Grant, Frank Knoefel, and Rafik Goubran. 2013. Bed occupancy measurements using under mattress pressure sensors for long term monitoring of community-dwelling older adults. In *2013 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. IEEE, 130–134.

[15] Sana Tonekaboni, Danny Eytan, and Anna Goldenberg. 2021. Unsupervised representation learning for time series with temporal neighborhood coding. *arXiv preprint arXiv:2106.00750* (2021).

[16] Ling Yang and Shenda Hong. 2022. Unsupervised time-series representation learning with iterative bilinear temporal-spectral fusion. In *International Conference on Machine Learning*. PMLR, 25038–25054.

[17] Zhihan Yue, Yujing Wang, Juanyong Duan, Tianmeng Yang, Congrui Huang, Yunhai Tong, and Bixiong Xu. 2022. Ts2vec: Towards universal representation of time series. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36. 8980–8987.

[18] Kexin Zhang, Qingsong Wen, Chaoli Zhang, Liang Sun, and Yong Liu. 2022. Time Series Anomaly Detection using Skip-Step Contrastive Predictive Coding. In *NeurIPS 2022 Workshop: Self-Supervised Learning-Theory and Practice*.

[19] Xiang Zhang, Ziyuan Zhao, Theodoros Tsiligkaridis, and Marinka Zitnik. 2022. Self-supervised contrastive pre-training for time series via time-frequency consistency. *Advances in Neural Information Processing Systems* 35 (2022), 3988–4003.

[20] Yunhao Zhang and Junchi Yan. 2022. Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting. In *The Eleventh International Conference on Learning Representations*.

[21] Yang Zhao, Peter Tu, and Ming-Ching Chang. 2019. Occupancy sensing and activity recognition with cameras and wireless sensors. In *Proceedings of the 2nd Workshop on Data Acquisition to Analysis*. 1–6.