WiStress: Contactless Stress Monitoring Using Wireless Signals

UNSOO HA*, Massachusetts Institute of Technology
SOHRAB MADANI*, University of Illinois at Urbana-Champaign
FADEL ADIB, Massachusetts Institute of Technology

Stress plays a critical role in our lives, impacting our productivity and our long-term physiological and psychological well-being. This has motivated the development of stress monitoring solutions to better understand stress, its impact on productivity and teamwork, and help users adapt their habits toward more sustainable stress levels. However, today's stress monitoring solutions remain obtrusive, requiring active user participation (e.g., self-reporting), interfering with people's daily activities, and often adding more burden to users looking to reduce their stress.

In this paper, we introduce WiStress, the first system that can passively monitor a user's stress levels by relying on wireless signals. WiStress does not require users to actively provide input or to wear any devices on their bodies. It operates by transmitting ultra-low-power wireless signals and measuring their reflections off the user's body. WiStress introduces two key innovations. First, it presents the first machine learning network that can accurately and robustly extract heartbeat intervals (IBI's) from wireless reflections without constraints on a user's daily activities. Second, it introduces a stress classification framework that combines the extracted heartbeats with other wirelessly captured stress-related features in order to infer a subject's stress level. We built a prototype of WiStress and tested it on 22 different subjects across different environments in both stress-induced and free-living conditions. Our results demonstrate that WiStress has high accuracy (84%-95%) in inferring a person's stress level in a fully-automated way, paving the way for ubiquitous sensing systems that can monitor stress and provide feedback to improve productivity, health, and well-being.

CCS Concepts: • **Human-centered computing** → *Ubiquitous and mobile computing systems and tools.*

Additional Key Words and Phrases: Stress, Wireless, Signal Processing, Machine Learning, Vital Signs, Heart Rate Variability

ACM Reference Format:
Unsoo Ha, Sohrab Madani, and Fadel Adib. 2021. WiStress: Contactless Stress Monitoring Using Wireless Signals. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 3, Article 103 (September 2021), 37 pages. https://doi.org/10.1145/3478121

1 INTRODUCTION

We all experience stress during our daily lives: while working on a deadline, preparing for an event, worrying about our children, or thinking about our futures. While moderate stress may boost our productivity [49, 116], chronic stress – or prolonged exposure to acute stress – has detrimental outcomes. It accelerates aging at the cellular level [92] and promotes an earlier onset of age-related diseases such as diabetes [82], Alzheimer's [81], and high blood pressure [29]. Clinical studies have also shown that chronic stress is correlated with increased risk of depression [114], fatigue [26], anxiety [95], and insomnia [119]. These studies demonstrate the importance

*Both authors contributed equally to this research.

Authors' addresses: Unsoo Ha, unsoo@mit.edu, Massachusetts Institute of Technology, Cambridge, Massachusetts; Sohrab Madani, smadani2@illinois.edu, University of Illinois at Urbana-Champaign, Urbana, Illinois; Fadel Adib, fadel@mit.edu, Massachusetts Institute of Technology, Cambridge, Massachusetts.

2474-9567/2021/9-ART103
https://doi.org/10.1145/3478121

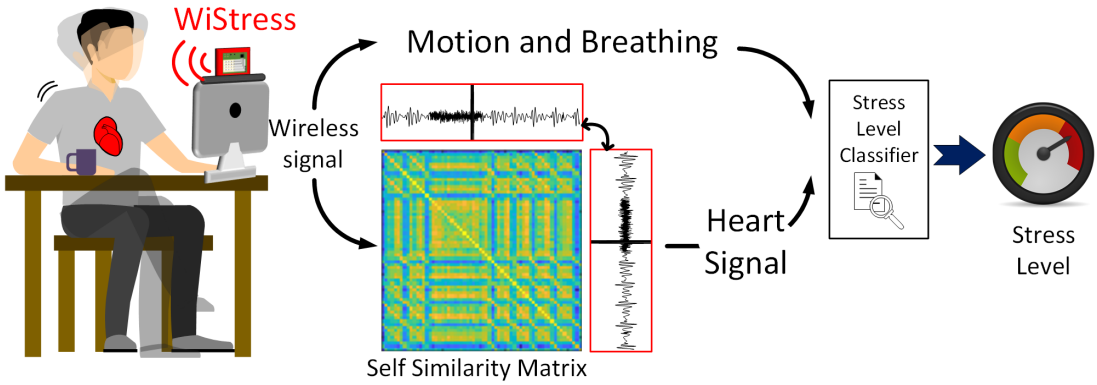


Fig. 1. Passive Stress Monitoring using Wireless Signals. WiStress measures wireless reflections off the person's body in order to sense stress-related biometrics and infer the person's stress level.

of continuous and fully-automated stress monitoring. Such monitoring can provide doctors with information to better manage patient conditions [14]; it can also help us modulate our own stress levels by incorporating meditation or adapting our daily activities to improve our well-being and productivity [95].

Unfortunately, today's stress monitoring solutions remain obtrusive and typically require active cooperation from users, thus interfering with their daily activities. The most standard approach relies on self-reporting, where users log their stress levels (e.g., low, medium, or high) in a journal or on a smartphone [74, 91], making this approach obtrusive [13, 63]. Another common approach for one-shot, acute stress measurement is to extract the cortisol hormone from saliva samples [33, 79, 105]. This approach is not suitable for long-term monitoring, as it relies on precise lab equipment, the subject's willingness to frequently provide saliva samples, and an expert's intervention to extract the hormone concentration.

Recognizing these limitations, researchers have looked into inferring stress levels from wearables that monitor stress-related biomarkers such as breathing, heart rate variability (HRV), and skin conductivity. These solutions fall into two major categories, which trade off comfort for accuracy [71, 87]. The first can accurately extract stress-related metrics but require bulky setups that can be burdensome for everyday use (e.g., chest bands that measure HRV [98] or shirts that measure muscle activity to infer stress [113]). The second category uses wristbands that can infer stress from HRV measurements. These approaches are more comfortable for everyday use but they have lower accuracy than chest bands due to the constant movement of a user's wrist [41, 45, 46], which has driven medical researchers to adopt protocols that use chest bands over wristbands for HRV measurements [69, 94]. Moreover, all of these approaches require users to actively participate and wear devices, which may be uncomfortable for various sections of the population [67].

An ideal stress monitoring solution would monitor a user in a completely passive way and without interfering with their daily activities. In this paper, we present WiStress, a system that can monitor a user's stress passively using wireless signals. As shown in Fig. 1, WiStress can be installed on a desk or near a couch to monitor a nearby user's stress levels. It works by continuously transmitting ultra-low-power wireless signals that reflect off the human body and capturing these reflections in order to infer the person's stress level. This solution is inspired by recent work that has demonstrated the potential to track a person's movements and vital signs using wireless signals [9, 10, 129]. Unlike this past work, it is the first to demonstrate the potential of using such signals to monitor a person's stress.

At the core of WiStress's design is a novel machine learning pipeline that can map wireless reflections to stress levels. The pipeline extracts three key stress-correlated biometrics from wireless reflections: breathing, heart-rate variability (HRV), and motion. Among these, HRV is particularly challenging because it requires sensing minute variations in the wireless reflections that arise from body movements triggered by heartbeats. Because heartbeat movements are very minute, they are easily masked by user movements as subtle as a shift in pose, nodding, shaking one's leg, or typing. As a result, unless the user is fully static, it is not possible to distinguish whether subtle changes in the wireless reflections are due to a heartbeat or due to a nod or an eye twitch [10, 128] let alone random noise, movements, or other users in the environment.

To overcome these challenges, our approach is to identify and leverage *temporally local self-similarities* in the wireless reflections and use them to zero in on a user's heartbeat. Specifically, rather than simply looking for subtle changes in the wireless reflections, WiStress looks for *similarities in these changes* over short windows of time. Since a user's heartbeats are repetitive and the heart rate varies gradually over time, this approach allows the network to zero in on the heartbeats. WiStress's method is particularly powerful because it can also eliminate subtle random movements (e.g., nods, typing) and quasi-random movements (e.g., shaking legs). To learn temporally local self-similarities, WiStress opportunistically captures wireless reflections over time and constructs a self-similarity matrix similar to the one shown in Fig. 1. As we describe in §3.2, the matrix compares time-shifted windows of reflections to each other, and feeds them to *WiStress-Net*, a network that can learn similarity features due to heartbeats while eliminating changes arising from extraneous movements.

WiStress builds on this fundamental technique to deliver a fully-automated system for passive stress monitoring. The system can automatically detect when a user is nearby and when they leave its sensing field. It incorporates techniques (described in §3.2) that enable it to automatically identify and segment the variations in reflections that arise from breathing and heart rate variability, and mitigate the impact of extraneous movements and interference. Furthermore, rather than entirely discard measurements with motion artifacts, we observe that WiStress can leverage motion to boost its stress classification accuracy. This is because certain body movements (e.g., frequently shaking one's leg or stretching their neck) are correlated with stress levels. In §3.3, we describe how WiStress's overall architecture enables extracting and selecting physiological and motion-based features to train its learning models to infer a user's stress level.

We built a prototype of WiStress using an off-the-shelf millimeter-wave sensing board (the TI IWR1443 module [5]). We tested our system on 22 subjects of different ages and genders, across different homes, and during many different daily activities, as well as specific tasks designed to induce stress. Throughout the experiments, subjects were free to move around, leaving and returning the radio range of the sensor; moreover, other people freely moved around in the background. To obtain ground-truth measurements during our long-term studies, we asked subjects to fill out a standardized NASA-TLX form [57] every 30 minutes.

Our results demonstrate that WiStress can passively and accurately classify among three standard levels of stress: low, moderate, and high. Its median accuracy is 90.7% when the models are tested and trained on the same person (while a random guess is 33.3%). Moreover, WiStress works correctly even when it is tested on people it has never been trained on (and in new environments); in such scenarios its median accuracy remains over 84%. We also demonstrate that WiStress can extract HRV's with very low error (median error < 4 ms) even when subjects are free to perform daily activities; in contrast, the error of state-of-the-art HRV extraction algorithms from wireless signals increases to around 50 ms (i.e., 10× that of WiStress) when subjects are allowed to perform daily activities, precluding the ability to use them for accurate and unobtrusive stress monitoring. Beyond obtaining spot-level stress checks, WiStress can track changes in a user's stress level over extended periods of time, paving way for future solutions that would allow users to monitor their stress levels and adapt their daily activities.

In summary, the paper makes the following contributions:

(1) We present WiStress, the first passive stress monitoring system that can infer a user's stress levels from wireless signals.
(2) We introduce WiStress-Net, a novel machine learning network that can extract heart rate variability (HRV) from wireless reflections without requiring users to remain static during the monitoring phase.
(3) We introduce a framework for extracting physiological and motion-based features from wireless signals for stress monitoring.
(4) We built a prototype of WiStress and tested it on 22 different human subjects in different homes. Our evaluation demonstrates that WiStress has high accuracy in extracting stress levels even when it has never seen the subject before.

Finally, it is worth noting that since WiStress relies on off-the-shelf millimeter wave radars (similar to those in the recently released Google Nest Hub [60]), it holds the potential to bring ubiquitous stress sensing to millions of users worldwide.

2 RELATED WORK

2.1 Stress Monitoring Systems

The past two decades have witnessed an increased interest in stress monitoring to prevent, diagnose, and even treat related diseases [21, 58, 68, 89]. The standard approach for stress monitoring relies on extracting cortisol from saliva samples or requiring users to complete questionnaires about their stress level. Because these methods are obtrusive and inconvenient for long-term stress monitoring, researchers have looked into a variety of alternative methods for inferring stress:

Wearable Devices. Recent studies have used a variety of wearables such as smart belts [87, 126], shirts [113], smart watches [16], and skin-attachable sensors [1, 71] for stress monitoring. These wearables typically infer stress by measuring biomarkers such as heart rate variability (HRV) [36, 85, 93], electrodermal activity [36, 59, 73, 84], and respiration [59, 80, 122, 124].

Although this body of work takes important steps towards less obtrusive stress monitoring, it has a number of some limitations. First, these methods still require attachment to the person's skin and/or body. As a result, certain population (e.g., children, elderly) may not be comfortable with wearing these devices for a long time, and a recent survey [71] reported that around 15 to 35 percent of the respondents would not be willing to continuously wear a stress monitoring device. Furthermore, these wearables typically exhibit a trade off between accuracy and comfort (more bulky devices are more accurate, but less comfortable), with certain devices requiring inconvenient skin preparation; for example, the BioStampCR [1] requires that any hair be removed from the area before the sensor is attached. Aside from the potential inconvenience of wearing these devices, the user must remember to recharge them on a daily basis [87].

WiStress shares the vision of less obtrusive monitoring with this past work, but in contrast to these prior systems, it does not require any contact with the user's body or any cooperation from the user (e.g., re-charging). More importantly, it opens up new use cases that require *location-specific* monitoring. For example, it can be incorporated into smart devices (like screens or kiosks) to infer when a user interacting with them is stressed and adapt to their stress level, or it can be installed in offices or classrooms to help monitor work stress of employees and/or students.[1]

Camera-based Methods. Camera-based methods have also been proposed to assess stress level by sensing visual cues such as a subject's head movements, blink rate, and pupil size variation [24]. Some methods also use videos to estimate the HRV of the subject in order to infer their stress level [48]. However, these methods involve

[1]Note that GPS-enabled wearable cannot enable location-specific monitoring in indoor environments because GPS accuracy is poor indoors [97].

capturing high resolution video of the subject to capture the minute changes corresponding to small variations in the user's face, which raises privacy concerns. Additionally, these methods are sensitive to lighting conditions and do not work in the dark [17]. In contrast, WiStress is much less privacy intrusive and works in the dark.

Smartphone-based Methods. Researches have also considered smartphones to monitor stress by analyzing user behavior [20, 23]. These analyses include speech analysis, location behavior, phone call and text patterns, and user personal traits. However, these studies have reported lower (70% or less) accuracy and reliability than bio-markers in inferring a user's stress, and raise privacy concerns for their users. Recent work [78] has also leveraged smartphones for photoplethysmogram-based HRV extraction, which is then used to assess the mental health of the subject. However, this approach shares the same challenges as contact-based methods since the user needs to put their finger directly on the smartphone camera and keep it there for the full duration of the experiment.

2.2 RF-Based Sensing

RF-based sensing is an emerging field which uses the reflections of RF (Radio Frequency) signals off the human body to track human movements, postures, and vital signs [10, 64, 121, 127, 130]. The closest to our work are past systems that measure heartbeats using RF signals. These can be divided into two main categories:

- The first category extracts the *average* heart rate from measurements collected over 1-2 minutes [8, 18, 31]. This category of systems can work with quasi-random user movements because it averages out these movements over the measurement period. However, it cannot extract individual heartbeats, which are of the order of 1 second; as a result, this category of systems cannot measure HRV at the level of individual heartbeats, which is required for stress monitoring [69].
- The second category can extract individual heartbeats and measure the HRV, but can only work in controlled lab environments. Specifically, these systems require users to sit or lie down in a specific position/orientation with respect to the sensing device, remain fully-static for the measurement duration [52, 101, 128], and often require users to hold their breath [101]. In the absence of such controlled lab setups, state-of-the-art systems (like EQ-Radio [128] and RF-SCG [52]) incur errors that are $10\times - 25\times$ larger than WiStress, making them incapable of inferring a user's stress level, as we demonstrate empirically in §5.1. As a result, using these systems for stress monitoring would require users to stop their daily activities in order to actively take a measurement. More importantly, they would make the user aware of being monitored, thus inducing the well-known subject-expectancy effect [22, 55] where the experimental setup biases the results.

WiStress builds on these prior systems to deliver the first fully-automated, passive system for monitoring stress by extracting HRV without imposing unnatural requirements on the user.

2.3 Self Similarity

Finally, our work is related to past work that uses the concept of self-similarity. Prior work has applied this concept to identify patterns in text [90], music [38, 39, 77], and videos [37, 40, 66, 108]. WiStress adapts this concept to the RF domain and is the first to apply it to the RF-based sensing.

3 WISTRESS'S DESIGN

WiStress is a passive stress monitoring system that relies on wireless signals. The system uses a wireless device that can sit on a user's desk or near their couch as shown in Fig. 1. The device continuously sends an ultra-low-power RF signal in the millimeter-wave band, and captures its reflection. It analyzes these reflections over time in order to detect the nearest user and infer their stress level.

WiStress operates in three main stages as shown in Fig. 2:

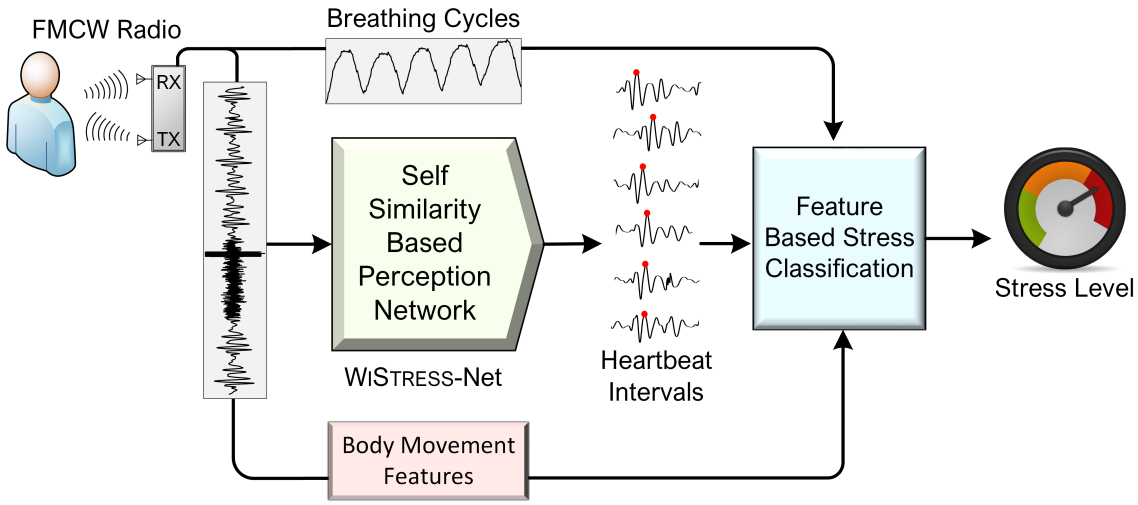


Fig. 2. Overview of WiStress's Design. The system uses a millimeter-wave FMCW radio to extract wireless reflections off the user's body. It filters these reflections to obtain time-domain signals that correspond to breathing cycles and micro-movements. The micro-movements are fed to a self-similarity-based perception network to extract individual heartbeats. Finally, the breathing cycles, heartbeats, and body movement features are fed into a stress classification network that infers a user's stress level.

(1) *Capturing the RF Reflection.* In the first stage, WiStress captures the reflection off a nearby user's body. This component can detect the presence of a user, determine when they leave and return to its sensing field (within approximately 3 m radius of the device), and eliminate the impact of other users and noise in the environment. It outputs a time-domain signal corresponding to the user's movements.
(2) *Self-similarity-based Perception Network.* WiStress's second component takes the time-domain signal corresponding the user's movements and outputs the user's heart rate variability. This component exploits a self-similarity matrix (SSM) to zero in on a user's heartbeats, and constructs a deep learning architecture that can robustly extract these heartbeats and eliminate extraneous movements.
(3) *Stress Classification Module.* WiStress's final component takes the heartbeats extracted from the perception network, selects and extracts stress-related features from the heartbeats as well as breathing cycles and movements, and uses the combined features to infer a user's stress level.

Next, we describe each component of our design in detail.

3.1 Capturing the Wireless Signal

The first step of WiStress's design is to capture the wireless reflection of a nearby user's movements. To do so, it transmits a low-power RF signal, measures its reflections, and filters them to zoom in on the nearby user. The main challenge in isolating the nearby user's reflections is that wireless signals not only reflect off that user's body, but also other objects in the environment, including furniture and other users.

To overcome these challenges, WiStress builds on past systems that employ radar techniques in order to isolate the user's reflection and eliminate those arising from other objects in the environment. We describe these techniques at a high-level and refer the interested user to prior work [8, 10] for more details. The techniques *isolate reflections arriving from each 3D location in the environment* by using a combination of Frequency-Modulated

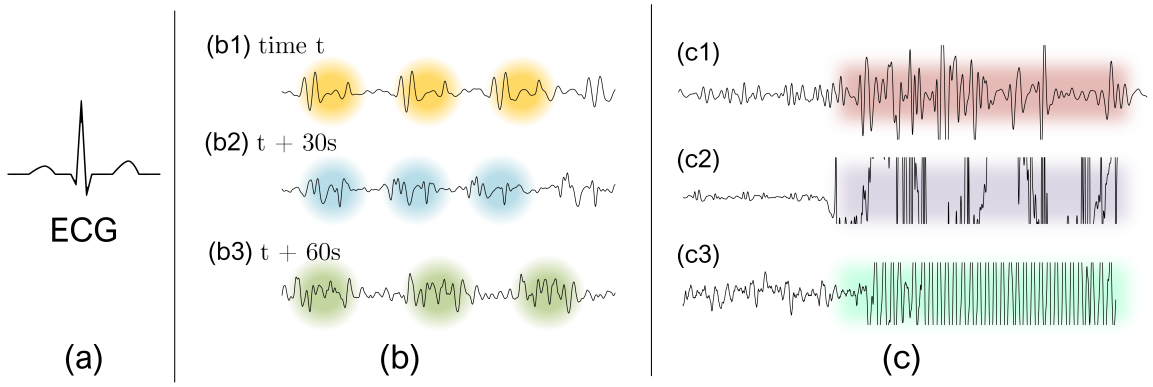


Fig. 3. Different patterns of heartbeat. (a): The ECG signal has a fixed heartbeat pattern. (b): heartbeats from wireless signals captured from a user in one experiment. (b1), (b2) and (b3) were captured 30 seconds apart. The color-coded highlights indicate different patterns that correspond to heartbeats. (c): Motion artifacts caused by user movements. (c1): eating food; (c2): moving back and forth in chair; (c3): shaking leg. Artifacts are different from each other, and cannot be predicted.

Continuous Wave (FMCW) radar and a 2D antenna array. Since different reflectors occupy different locations in 3D space, WiStress uses these techniques to isolate different reflectors into separate buckets. Subsequently, it eliminates all buckets with static reflections (e.g., furniture, walls) and identifies the nearest bucket that corresponds to a moving user.[2]

Once the system identifies the bucket that corresponds to a user, it extracts the phase of the wireless signal in that bucket. The phase captures any small changes in distance between the user's body and the sensor's antennas. Since both breathing and heartbeats are associated with variations in distance (due to chest movement or micro-vibrations on the surface of the body), the obtained time-domain signal encodes movements corresponding to these vital signs. Mathematically, the phase is given by the following equation [117]:

$$\phi(t) = 2\pi \frac{d(t)}{\lambda}, \tag{1}$$

where $\phi(t)$ is the phase of the received signal, λ is the wavelength of the signal, and $d(t)$ denotes the the distance over time between the device and the human body. Since heartbeats appear in $d(t)$ by producing small mechanical vibrations on the user's chest area, WiStress can sense these vibrations through Eq. 1 by extracting the phase $\phi(t)$ from the received signal and applying a double differentiator filter similar to prior approaches [52, 128].

3.2 Self-Similarity-Based Perception Network

Now that WiStress has isolated the reflections from a nearby user's body, it proceeds to extracting individual heartbeats from these reflections. To do so, it passes the time-domain signals into a self-similarity-based perception network, *WiStress-Net*, which sits at the core of WiStress's design. In this section, we first explain the motivation and rationale for the self-similarity network. Then, we describe how WiStress designs this network to extract individual heartbeats using a hybrid architecture of signal processing and deep learning.

3.2.1 The Challenge in Passively Extracting the HRV. Passively extracting the heartbeats from a user's wireless reflection is very challenging. Unlike contact-based physiological biometrics like ECG, which produce a known

[2]The system is sensitive enough to pick up reflections that are varying due to other moving objects, e.g., fan or small pet. In such scenarios, it can eliminate these buckets using the technique described in §3.2.

waveform as depicted in Fig. 3(a), heartbeat patterns in wireless signals are different each time, and are not known a priori. Even within the same experiment and for the same person, this pattern can change. Fig. 3 (b) shows the heartbeat patterns of a user during one experiment, captured thirty seconds apart by our wireless device. The different colors in the figure highlight that each of (b1), (b2) and (b3) have their own heartbeat morphology (shape), even though they come from the same user. The reason is that even a small shift in the user's pose during the experiment alters their wireless reflections and changes the captured micro-vibrations arising from their heartbeats.

The second challenge in extracting heartbeats arises from a user's unpredictable body movements. Fig. 3 (c) shows three examples of the captured wireless reflections in everyday scenarios. In the figure, (c1), (c2) and (c3) correspond to the user eating food, moving back and forth in their chair, and shaking their legs respectively. Let us consider (c1), where the user is eating food. Different acts while eating such as placing the food in mouth, chewing, and swallowing each give rise to different movements in parts of the body. These movements contaminate the captured reflections, easily masking the user's heartbeats. Moreover, these motion artifacts are difficult to filter out since (unlike breathing) they do not occur at a predefined set of frequencies. As highlighted in the other examples (c2) and (c3), moving in chair and shaking a leg, each produce their own artifacts, with different amplitudes.

All of these examples motivate the need for a technique that can (a) reject unpredictable motion artifacts that corrupt the user's wireless reflections, while (b) being able to quickly adapt to the continuously changing morphology of heartbeats in the user's wireless reflections.

3.2.2 Exploiting Self-Similarity. Our idea to overcome the above challenges is to exploit *temporally local self-similarities* in the captured phase signal. Let us understand this idea in the context of Fig. 3. Notice that in Fig. 3(b1)-(b3), although the heartbeat morphology may change over time, each morphology locally repeats multiple times. Thus, we can exploit this local self-similarity in order to identify the heartbeats. In particular, recall that our goal is to extract the length of individual heartbeats (HRV). Thus, if we can search for and identify local self-similarities, we can use them to extract the heartbeats.

To translate this idea into a practical system, we construct a self-similarity matrix (SSM), as shown in Fig. 4(a). The matrix takes as input two copies of the reflected wireless signal (shown as black time-domain signals to the top and rotated to the left-side of the figure) and computes a similarity metric between them. The figure is in the form of a heatmap, where yellow indicates high similarity and blue indicates low similarity. Let us take the two segments highlighted by the horizontal and vertical red dashed lines. These segments contain very similar patterns. As a result, their corresponding locations in the SSM (the square region at the intersection the four dashed lines) have a high value (yellow-ish color). Since the entire input signal has heartbeat signals similar to each other, we see the same highlighted square pattern repeat itself across the SSM. This is why we say the SSM encodes temporally local self-similarities in the signal. We can therefore use the SSM to extract repeated patterns from short periods of time.

Fig 4(b) shows the SSM obtained from another time segment of the user's reflections. Similar to Fig. 4(a), the input signal in Fig. 4(b) contains repeated heartbeat patterns, except that the repeated patterns in (b) are different from those in (a). Despite their different patterns, the corresponding SSMs for Fig. 4(a) and (b) are very similar and capture the periodicity of the repeated signals. This indicates that SSM does not depend on the particular morphology in the signal, but rather it depends on similarities between the those morphologies. Therefore, by looking at morphologies the SSM of the signal, we can circumvent the problem of identifying the shapes of the patterns in the signal and instead shift our focus on whether those morphologies are repeated.

Next, let us see how WiStress can use the SSM to deal with motion artifacts. To demonstrate this, we take one of our previous examples, specifically, the experiment with the user shaking their leg in Fig. 3 (c3), where the user starts to shake their leg during the measurement. Fig. 5 (a) shows this signal along with its corresponding SSM. We make two observations:

- First, in the portions of the signal where there are heartbeats, the corresponding regions in the SSM still show patterns similar to those in Fig. 4.
- Second, when the user starts shaking their leg, these patterns disappear in the SSM, and we no longer see clear pattern, even though it seems shaking leg is a repetitive motion. This is because we use a feature extraction network before computing our SSM (explained in §3.2.3) that only extracts features related to heart signals, and rejects other motions such as shaking legs, even if they include repeating patterns.

This example demonstrates the SSM is able to mask out the motion artifacts and noisy parts of the signal while still encoding the similarities between existing heartbeat patterns.

Finally, let us consider another example of distorted wireless reflections, whose SSM is shown in Fig. 5 (b). In this example, the user is constantly moving their limbs in different directions. Since the user is moving during the entire measurement of this signal, their heartbeats are completely masked by motion artifacts. Looking at the SSM, we do not observe any patterns like those that appeared in the previous figures (aside from the yellow diagonal which corresponds to the self-correlation). This shows that the SSM may be used to reject regions that do not include any heartbeats, including times when the user is moving so much that the heartbeats are masked, or is not close enough to the device to be detected by the system.

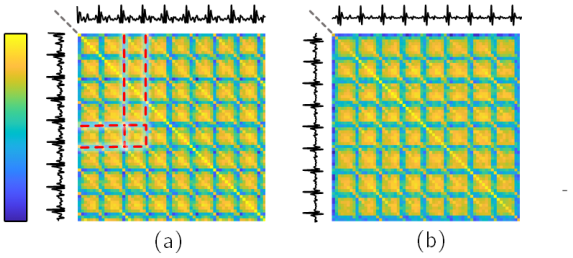


Fig. 4. Self similarity matrix (SSM) in the presence of high quality wireless signals that contain heartbeat. Different segments of the input signals (drawn in black) are compared with each other in terms of similarity to obtain the self similarity matrix. Although (a) and (b) have different heartbeat patterns, their corresponding SSMs look almost the same.

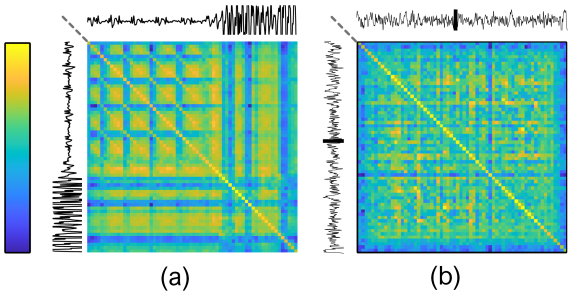


Fig. 5. Self similarity matrix (SSM) when the wireless signals are corrupted by motion. In (a), the user starts shaking their leg in the middle of measurement, similar to Fig. 3 (c3). While we can still see the heartbeat patterns, the patterns disappear when user starts shaking their leg. In (b), the user is moving constantly and the signal and does not contain heartbeats. Thus, the SSM does not show any clear patterns.

3.2.3 WiStress-Net. The above discussion demonstrates that by employing an SSM, WiStress can capture local self-similarities in wireless reflections. Next, we describe how WiStress uses this SSM in order to extract individual heartbeats.

Fig. 6 shows the overall pipeline of WiStress-Net, the perception network employed by our system to isolate individual heartbeats. The network consists of two sub-networks: (1) an SSM computation network, which takes the wireless signal as input and outputs the SSM, and (2) a heartbeat extraction network, which takes the computed SSM as input and extracts the precise temporal location of each heartbeat inside the original input signal.

Before we describe these two sub-networks, let us formalize the definition of a self-similarity matrix.

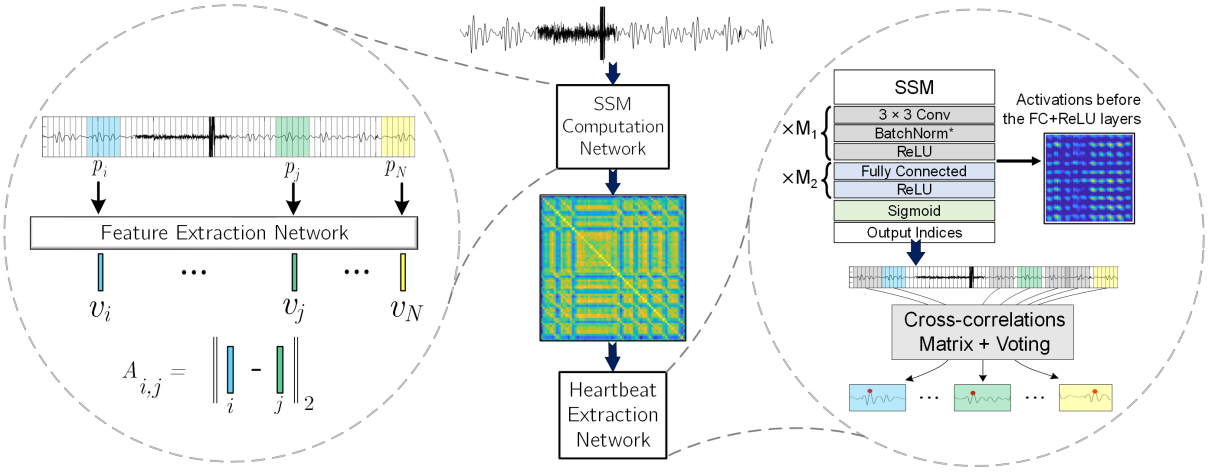


Fig. 6. WiStress-Net's pipeline. Three color-coded segments (blue, green and yellow respectively) are highlighted throughout pipeline. Middle: The main pipeline. The SSM computation network is used to compute the self similarity matrix, which is in turn fed into the heartbeat extraction network that extracts individual heartbeats. Left: The SSM computation network. Right: the heartbeat extraction network.

Definition of the Self-Similarity Matrix (SSM). Given a set of segments $\{p_1, \cdots, p_N\} \in \mathbb{R}^N$ and a similarity function $f : \mathbb{R}^N \times \mathbb{R}^N \rightarrow \mathbb{R}$, the SSM is defined as a $N \times N$ matrix A, where $A_{i,j} = f(p_i, p_j)$.

SSM Computation Network. The first component of WiStress-Net is a network that computes the SSM. The first step in computing the SSM is to choose the self-similarity function f. Unfortunately, choosing simple functions like euclidean distance applied directly on the input signal will not capture the similarities between heartbeat patterns well [37, 123]. Instead, WiStress learns this function through the **feature extraction network** depicted to the left side in Fig. 6.

The first step of the network divides the input signal into smaller segments, $p_1, \cdots, p_N$. Next, each signal segment is independently passed through the feature extraction network. Figure 6 (left) shows the high-level process of passing all the segments through the network. As shown in the figure, once all feature vectors $v_1, \cdots, v_N$ are computed, the ij^{th} element of the SSM matrix $A_{i,j}$ is computed as the euclidean distance between v_i and v_j. Hence, the feature extraction network, together with the euclidean distance define the similarity function f.

In designing this network, we experimented with different lengths and partitions of the input signal. We found that two design criteria helped in improving the robustness and generalizability of the network:

- First, the input signal is segmented in a way to have overlapping regions between adjacent segments. This is because overlapping regions help share information between segments. All the possible starting points for a typical signal are depicted in the figure as vertical lines. Specifically, two adjacent starting points are 100 ms apart.
- The length of the segments are chosen such that their duration is less than the minimum heartbeat period but large enough to capture salient features (e.g., peaks or valleys) in the heartbeat morphology. Satisfying these two constraints helps the subsequent heartbeat extraction network in classifying each segment as having zero or one heartbeat. Empirically, we found that 400 ms was a good segment duration. This means that the overlap between adjacent segments is 300 ms.

Heartbeat Extraction Network. The second component of WiStress-Net is the heartbeat extraction network. This network takes the SSM as input and outputs individual heartbeat intervals in the signal. To this end, WiStress uses a 2D convolutional neural network depicted in Fig. 6 (right) that acts directly on the self similarity matrix. The goal of this network is to classify each of the N segments involved in the SSM into one of the two simple groups: ones that contain a heartbeat, and ones that do not contain a heartbeat. The details of the neural network, along with examples of intermediate representations in between layers are shown in the figure. The final output of the network is the set of indices of the patterns that are predicted to be a heartbeat. As can be seen from the figure, there are 7 such identified heartbeat patterns for the given signal, three such segments are highlighted using the colors blue, green, and yellow.

Next, for this group of identified heartbeat patterns, WiStress zooms in on each of the coarse time segments to extract fine-grained beat-to-beat intervals. To this end, WiStress creates a new, smaller matrix (for example a 7 by 7 matrix in Fig. 6). Row i of the matrix denotes high resolution time-shifts between i and all the other patterns that maximizes their mutual similarity. Therefore, each segment i casts a vote on how much segment j should shift in time to maximize their similarity. Then, for each segment i, these votes are resolved by taking the median of all the candidate shifts. Finally, by compensating for these mutual time-shifts, WiStress finds the inter-beat-intervals (IBIs)[3] between consecutive heartbeat patterns.

Interestingly, we found that for the high-resolution heartbeat intervals, the cross-correlation between the raw signal segments (p_i) work better than the extracted features (v_i). We believe that the reason behind this is twofold; first, the feature extraction network includes a global max-pooling over the temporal information which is detrimental to finding the find-grained temporal location of the heartbeat. Second, the similar heartbeat patterns have already been isolated by the preceding network; in other words, if the network is trained properly, the identified patterns should be very similar, like each row in Fig. 3 (middle).

Few additional points are worth noting:

- One might wonder whether a single end-to-end convolutional network may achieve the same performance as WiStress-Net in extracting individual heartbeats. The reason why such a network would be less effective is that it cannot capture the temporal self-similarities which are reflected in the SSM. Indeed, as we demonstrate in §5.1, prior systems that applied such an approach to extract HRV [52] achieve much lower accuracy than WiStress in the presence of noise and unpredictable movements.
- Before feeding the signal into the SSM computation network, WiStress applies a band-pass filter to reject the impact of breathing, similar to past proposals for vital sign sensing [52, 128].
- Aside from extracting heartbeats, WiStress also extracts breathing signals. After discarding segments with extraneous movements using WiStress-Net, it identifies the segments in the original input signal (from Eq. 1) that are free from motion artifacts. Subsequently, it applies a low-pass filter on these segments in order to extract the user's breathing signal as shown in Fig. 2.

3.3 Stress Classification Module

WiStress's third and final component is a stress classification module, which takes the extracted IBI measurements as input, combines them with other features obtained from wireless reflections, and outputs the user's stress level. A standard challenge with IBI-based stress monitoring using contact-based methods (e.g., ECG, PPG) is that motion artifacts contaminate the signal, and the contaminated segments must be discarded prior to feeding them to the stress classifier [27, 51, 86, 87]. However, adopting the same approach in WiStress is undesirable for multiple reasons as we explain below.

[3]IBIs are HRV values at the granularity of one heartbeat.

In the context of wireless sensing, simply discarding all segments with motion artifacts will negatively impact WiStress's overall performance. This is because wireless signals are much more sensitive to the human's motion than contact-based modalities. Specifically, the captured wireless reflection is affected by various kinds of body movements within the antennas' field-of-view (FoV), whereas wearables are affected only by the local body motions (e.g., moving the right hand while wearing a smartwatch on the left hand will not introduce noise). Moreover, since the signal reflections representing the mechanical movement of the heart are so weak, they are easily buried in other motion signals. Therefore, in a scenario where a person is moving, in order to receover accurate IBIs, we need to discard those contaminated regions more often. Unfortunately, the discontinuity and missing values will distort the features and mislead the estimation (as we will demonstrate in §3.3.2), which would in turn reduce the stress classification accuracy.

To overcome this challenge, WiStress introduces three new techniques into its stress classification module. First, while motion artifacts are typically considered harmful in standard contact-based modalities, we observe that motion itself contains meaningful information that can help in stress monitoring. For example, people under high stress typically exhibit specific body language such as frequently changing their body posture or shaking their foot/hand more often [11, 12, 47]. Given its field of view, WiStress can sense these movements and incorporate them into its stress classification module.[4] Second, while harnessing motion patterns can improve the accuracy of classification, WiStress still needs to discard the corresponding time segments from its IBI-extraction network to avoid errors in IBI estimation, but such discarding results in a sparse time series of IBI measurements. The second technique in WiStress's stress classification module enables it to account for a sparse time series of IBI measurements. Third, aside from IBI and motion patterns, WiStress also extracts respiration signals from the wireless reflections, and uses them to enhance its stress classification accuracy. As we will demonstrate empirically in §5.1, all three techniques meaningfully contribute to WiStress's overall classification accuracy.

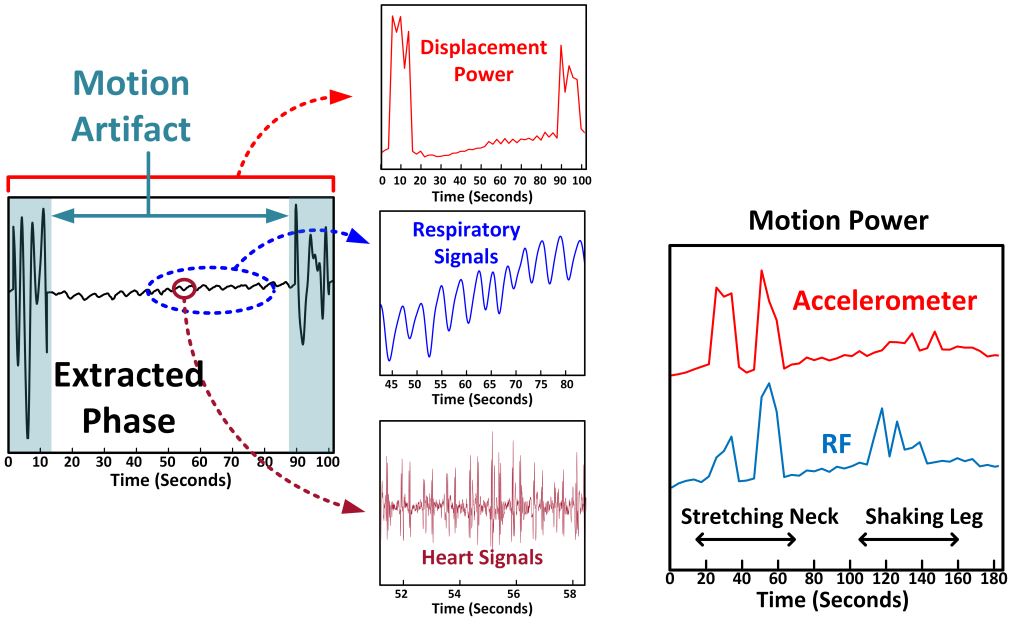
Before describing the above techniques in detail, we describe an experimental trial demonstrating the three types of features described above. Recall from §3.1 that WiStress captures the phase of the wireless reflection of the human body. Fig. 7 (a) plots variations in phase over time during an a short experiment that lasts for 100 seconds. Notice that in the beginning and the end of the experiment, the phase has significant variations due to motion artifacts. This is because for the first and last 15 seconds, the subject waved his hands. The figure shows how WiStress can extract three different time-series from the phase signal, representing body movements, breathing, and heartbeats. The figure also shows how WiStress can identify segments when the user is quasi-static (15-85 s) to extract physiological features as well as extract motion-related features from the IBI-discarded regions (0-15 s, 85-100 s).

The rest of this section details the techniques in WiStress's the stress classification module. The overall flow of the stress classification module is shown in Fig. 8. Below, we describe the different signal processing techniques applied to each of the three processing chains.

3.3.1 Body Movement Features. First, we describe how WiStress extracts motion-related features from the RF reflections. To demonstrate that WiStress can sense stress-related motion features, we perform an experiment where the user is asked to perform certain movements. In the experiment, the subject sat about 3 ft away from WiStress's antenna. To compare our system to a contact-based sensing approach, we also asked the subject to place an accelerometer on their chest, with the goal of detecting the body movements.

Fig.7(b) plots the output of the accelerometer and that of the RF sensed displacement. The red line represents the computed displacement power from an accelerometer (which is a well-known behavioral feature [42, 47, 125]),

[4]Note that motion alone is typically not a sufficient feature to achieve high accuracy in stress classification. In particular, the subject may be moving due to a non-stress related reason. Hence, WiStress uses motion as a contributing (rather than sole) feature in its stress classification task.



(a) Phase variation due to different factors.

(b) Displacement power of accelerometer and RF.

Fig. 7. Extracting Physiological and Motion-based Features for Stress Classification. (a) The figure shows the variations in phase due to breathing, heartbeats, and body movements. For the first and last 15 seconds, the subject waved his hands. The feature extraction module can acquire features from three different domains. (b) The figure shows the computed power of displacement from accelerometer (red) and RF signals (blue). During the experiment, the subject stretched his neck (20-70s) and shook his leg (110-160s). The RF modality can sense the motion in both cases, but the accelerometer cannot detect the leg's movement.

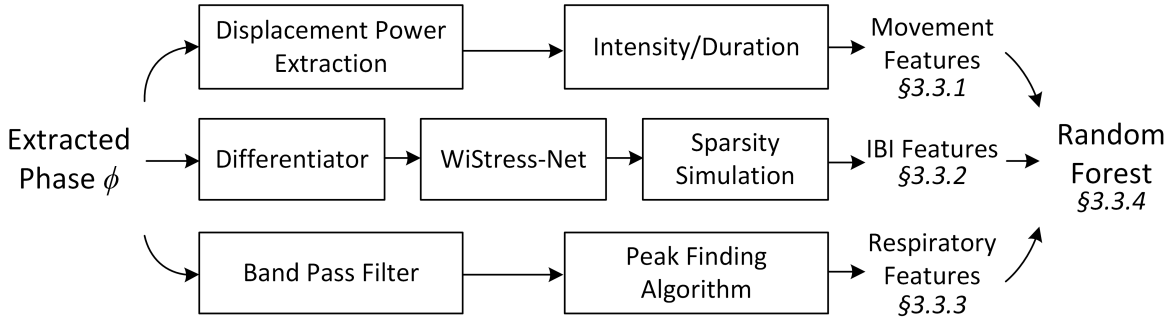


Fig. 8. Classification flow. WiStress uses three feature domains for stress level estimation. Each feature is extracted from the phase signal, processed through different pipelines, then fed into a random forest classifier.

and the blue line represents the magnitude of displacement computed from the RF reflection (described below). During the experiment, the subject stretched his neck (20-70s) and shook his leg (110-160s). Notice that while the

accelerometer can detect only the neck motion, the RF modality can sense both motions. This is because the accelerometer placed far from the leg and cannot sense the tiny motion.

Below, we detail the different features extracted from the captured RF phase:

(1) *Movement intensity*: The Movement Intensity (MI) feature is the power of displacement of a unit window (3 min). For a given unit window, we computed the feature that is the sum of the displacement power between two sample points.

$$MI(W_i) = \sum_{n \in W_i} (\phi[n+1] - \phi[n])^2 \tag{2}$$

where $MI(W_i)$ denotes the movement intensity feature of i^{th} unit window W_i, and $\phi[n]$ denotes the n^{th} sample point of the extracted phase signal.

(2) *Number of high activity occurrences*: The number of high activity occurrences represents how often large motion is detected in a unit window.

$$NoH(W_i) = \textbf{card}(\{j \,|MI(W_j) > P_{th}\}) \tag{3}$$

where $NoH(W_i)$ denotes the number of high activity occurrence of i^{th} unit window W_i, **card** denotes the number of element in a set, P_{th} denotes the threshold power for high activity detection[5], and W_j denotes the j^{th} sliding window in a unit window W_i.

(3) *Mean intensity of high activity*: It represents how large the detected high activities is in a unit window.

$$MiH(W_i) = \sum_{MI(W_j) \in S_{W_i}} MI(W_j)/\textbf{card}(S_{W_i}), \quad \text{where } S_{W_i} = \{MI(W_j) \big| MI(W_j) > P_{th}\} \tag{4}$$

where $MiH(W_i)$ denotes the mean intensity of high activity in i^{th} unit window W_i, and S_{W_i} denotes a set of power of j^{th} sliding window W_j in W_i.

Finally, it is worth noting that WiStress's approach is agnostic to the specific type of body movement (e.g., shaking leg or stretching neck). In particular, rather than trying to discover the body movement, it automatically leverages the above movement-agnostic features in its classifier (described in §3.3.4).

3.3.2 IBI Features. Next, we describe how WiStress selects and extracts features from the IBI measurements obtained in §3.2 and how it makes its classifier robust to sparsity in the IBI time series due to discarding segments with motion contamination. Since the WiStress-Net module is able to extract IBIs, it can compute IBI features commonly-used in stress monitoring [7, 70, 96]. These features are typically classified into temporal, frequency, and non-linear domains:

(1) Temporal Domain: *Mean* of IBIs, Standard deviation of IBIs (*SDRR*), Root Mean Square (*RMSSD*) and standard deviation (*SDSD*) of IBIs' successive differences, The percentage of the number of successive IBIs varying more than 50ms from the previous interval (*pNN50*)
(2) Frequency Domain: High Frequency Power (*HF*)
(3) Non-linear Domain: Poincaré anaysis (*SD2/SD1*)

A key question in selecting features is whether WiStress should leverage all the above features in training its classifier, similar to contact-based stress monitoring methods. Recall that one main difference here is that WiStress needs to discard a significant number of segments due to motion contamination, which leads to sparsity in IBI time series. However such sparsity may impact the accuracy of some of the above IBI features (which would in turn reduce the overall accuracy of stress classification).

[5]In our experiment, P_{th} = 13.

To investigate the effect of sparsity on each of the features, we perform experiments to emulate the discarding of different portion of the IBI time series, and assess the impact of such sparsity on the accuracy of the each of the computed features. We used 50-minute IBI series from five static subjects. The IBI features were extracted from 3-minute segments with a 20-second sliding window. Since motion is continuous, we expect its impact to be be on a consecutive chunk of the time series. To simulate such impact, we chose random chunks of different sizes and removed their IBI estimates from the time series. We repeated this simulation ten times for three different scenarios (discarded percentage: 30%, 50%, 70%) and compared the feature values to those computed using a dense time series (i.e., where 0% of the IBI's were discarded). Mathematically, we computed the error as follows:

$$\text{Error} = \frac{\left|\text{Computed Feature}(n\%) - \text{Ground Truth}\right|}{\text{Ground Truth}} \times 100 \tag{5}$$

where Ground Truth($n\%$) denotes the feature values when $n\%$ of the segment is discarded.

Table 1. Error in IBI-based Feature vs Discarded Portion of the Time Series. The table compares the error of IBI features as a function of the discarded portion.

IBI Features	Discarded Portion		
	30%	**50%**	**70%**
Mean	1.4%	1.8%	2.1%
SDRR	4.8%	6.8%	9.3%
RMSSD	3.9%	5.4%	8.6%
SDSD	11.1%	13.8%	19.3%
pNN50	3.3%	8.9%	13.5%
LF	8.9%	13.8%	24.4%
HF	6.2%	8.9%	16.6%
LF/HF	11.3%	23.8%	31.7%
SD2/SD1	3.8%	6.2%	8.8%

Table. 1 lists the error in computing each of the features as a function of the discarded portion. Interestingly, the temporal and non-linear features are less affected by the discarded regions. In contrast, the frequency features are sensitive to the changes. This is expected because missing data will change the phase of the overall frequency bin and make the frequency resolution worse.

Based on these results, we decided not to use the LF and LF/HF features in WiStress's overall classification module. Moreover, to make the classifier robust to missing/discarded IBI's, we added sparsity simulation module to the training regime that augments the IBI features similar to what we did in the simulation for Table 1.[6] Specifically, it reproduces features given the missing IBI assumption and repeats the augmentation ten times for each window. We didn't remove any IBI's if the original series has more than 70% of missing region in a unit window. By doing this, our classification model can learn the patterns and variations of features for real-world measurements.

3.3.3 Respiration Features. Aside from motion and IBI, WiStress also extracts respiration features from the RF reflection. Extracting respiration is motivated by prior work that has demonstrated that changes in the speed and depth of respiration are correlated with stress levels [59, 107, 122, 124].

[6]This is the white gaussian noise replacement augmentation technique described in §4.5.

To extract respiration signals, WiStress adopts similar techniques to prior wireless sensing methods for breath monitoring [10, 52]. In particular, as a user breathes, their chest expands and contracts, changing the distance to WiStress's antennas and impacting the captured wireless signals as shown in the middle plot of Fig. 7. Since respiration signals have lower frequency and higher amplitude than those of heart signals, they can be extracted from the captured phase by applying a standard band-pass filter.[7] Hence, WiStress applies a bandpass filter after identifying the regions without motion artifacts then it uses a standard peak-finding algorithm to identify the local maxima and minima [83], which correspond to the inhale and exhale process. Subsequently, it computes the depth (peak-valley) as well as the *Mean, SDRR, RMSSD* and *SDSD* of respiration.

3.3.4 Stress Classifier. Once WiStress obtains features from motion, IBI, and breathing, it can proceed to the final stage of stress classification. To build our classifier, we adopted a similar approach to prior stress monitoring systems which relied on *random forests*. The random forest is an ensemble algorithm which learns features by constructing multiple decision trees [62] and has been demonstrated to achieve high accuracy in sterss monitoring [27, 50]. In contrast to prior methods which only relied on IBI features, we feed our classifier all three types of features described above. The algorithm selects random subsets from the extracted features to create a large number of decision trees. Each decision tree makes classification output based on the subsets from the respiratory, body movement, and IBI features and the final output is determined by taking the majority vote algorithm from all the decision trees. The performance of the resulting classifier is evaluated in §5.1.

4 IMPLEMENTATION

4.1 Hardware

WiStress combines a millimeter-wave sensor with a real-time data-capture adapter for radar sensing. The millimeter-wave sensor is an IWR1443BOOST board [5]. It transmits an FMCW radar signal whose center frequency is 77 GHz and bandwidth is 4 GHz. The board also has two linear arrays for beamforming: horizontal (with 3-dB beam-width of ±28°) and vertical/elevation (with 3-dB beam-width of ±14°), implemented as a 3-switched-transmitter and 4-receiver system. The captured data is sent to a host PC using a mmWave Studio software developed by TI [6].

4.2 Software

The pre-processing of millimeter-wave data, including beam-forming and filtering was done in MATLAB 2020b. Other components including training and testing the networks, and data augmentation were done in Python 3.8. The parameters of IWR1443BOOST described in §4.1 were configured using mmWave Studio [6]. In order to design a dedicated stress elicitation software, we used PsyToolkit software [111, 112] which is frequently used for academic studies of cognitive science.

4.3 Baselines

To compare WiStress with previous designs, we implemented three state-of-the-art baselines as explained below. In all of the baselines, we removed a part of WiStress's pipeline and replaced it with algorithms used in previous work.

RF-SCG [52]. This work leverages a template learning method to extract IBIs from wireless signals. For this baseline, we implemented the IBI extraction algorithm (as described in [52]) and used it in WiStress's pipeline instead of WiStress-Net. The rest of WiStress's pipeline remained unchanged.

[7] The passband spans $[0.05\ 0.5\ Hz]$.

EQ-Radio [128]. This work also introduces an algorithm to extract IBIs from wireless signals. For this baseline, we implemented the IBI extraction algorithm (as described in [128]) and used it in WiStress's pipeline instead of WiStress-Net.

PPG. In this baseline, instead of the millimeter-wave modality, the IBIs were measured by placing the PPG sensor [4] on a user's fingertip (or toe, as described in the respective evaluation section). The IBIs from the PPG signal were computed using a simple peak detection algorithm, then manually reviewed by three people. Once the IBIs were extracted, we fed them as we would through WiStress's classification pipeline. Since breathing cannot be typically extracted from PPG measurements, we augmented the IBIs extracted from PPG with the breathing signals obtained from WiStress in our evaluation of this baseline (whenever needed).

The PPG setup we implemented employs a figertip Photoplethysmogram/PPG (AFE4400SPO2EVM [4]) which supports low-noise (13pA_{rms} noise current) and high-resolution (22-bit ADC) sensor front-ends. The millimeter-wave clock was shared with the PPG board for the sampling synchronization. To do so, XIN pin of the PPG board and MCUCLKOUT pin of the millimter-wave board were connected. The sampling rate was set to 500 samples per second for both boards, similar to WiStress's sampling rate.

ECG. For the WiStress-Net's generalizability test in §5.4.2, we employed a Biopotential Demonstration Kit (ADS1299 [3]) which has low-noise (1μVpp) and high-resolution (24-bit ADC) sensor front-ends. We used wet-electrodes for the measurements.

4.4 Training The Networks

Dataset. The two networks used for heartbeat feature extraction and heartbeat extraction were trained end-to-end using our captured millimeter-wave data and a publicly available SCG dataset [43]. For the millimeter-wave data, 4 people manually and independently annotated the heartbeats, which we aggregated and used as ground-truth. In total, our training data contained 50 hours of 1D heart recordings. The set of subjects whose heartbeats were used for heartbeat detection training were disjoint from the set of subjects on which we evaluated on system.

Network Parameters. The neural networks were written in PyTorch. The cross-entropy loss and an ADAM optimizer with $(\beta_1, \beta_2) = (0.9, 0.999)$ was used to optimize the networks. We started the training with a learning rate of $1e-4$ and reduced the learning rate with a factor of 0.3 whenever the validation loss plateaued for more than 5 consecutive epochs. During training, we used a batch size of 8 and set the values $M_1 = 6$ and $M_2 = 2$ for the heartbeat extraction network in Fig. 6 (right). We fixed length of the network's input signal to be 6.7 seconds (3550 samples), and set the duration of each smaller segment (p_i in section 3.2) to be 400 milliseconds. The size of the self similarity Matrix was set to be 64×64.

Random Forest. We used a random forest [76] library from sklearn [2] with $n\_estimators = 500$, $criterion =$ "$gini$", $max\_depth = 3$.

Feature Extraction Network (from §3.2.3). The network is comprised of five layers each with 1, 8, 16, 32 and 64 channels. Each layer has a 1D convolution with ReLU activation and BatchNorm. At the end, a global max-pooling is applied to all channels to get 64 scalars as a feature vector v. Note that the max-pooling is performed over the temporal dimension that removes the dependence of the features on time.[8] This is particularly important since we care about similar patterns within each segment, not whether they happen at the end or the beginning of the segment.

[8]This is similar to how the max() function returns the same value if we shift the signal in time.

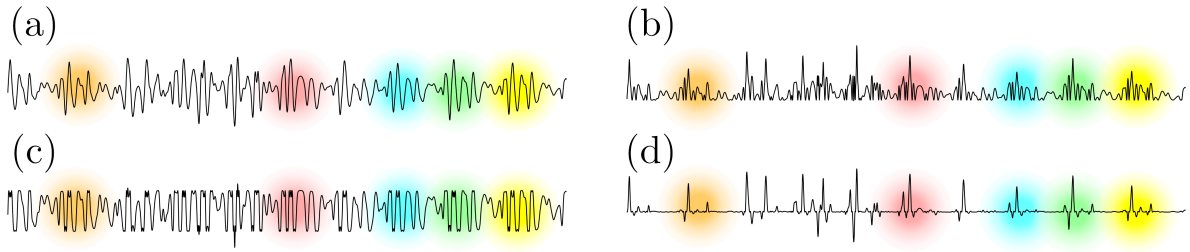


Fig. 9. Examples of non-linear data augmentation via random polynomials used in WiStress. (a) is the original captured signal. Each of (b), (c) and (d) are copies of (a) with random polynomials applied. A group of color-coded patterns across the signals are highlighted, indicating that similarities between patterns are preserved under the effect of polynomials.

4.5 Data Augmentation

To help train our networks and prevent overfitting, our training regime incorporates a variety of standard data augmentation techniques from the machine learning community. Data augmentation improves the robustness of learning models to noise and to unseen variations in the measurement dataset. We implemented online data augmentation [30] with 100 epochs.[9] In each epoch, every data point (in the context of WiStress-Net, each data point is a 6.7-second time series signal, as described above) is modified prior to feeding it to the network. Below, we enumerate the augmentation techniques we chose and explain the rationale for each of them.[10]

- *Time-shifting:* This augmentation technique helps the network to be *shift invariant*, i.e., robust to temporal shifts in the input signal. To implement this technique, the input signal is shifted by a randomly chosen number of samples between 0 and 500 samples (corresponding to 1 second, i.e., a standard heartbeat).
- *Linear Time Expansion/Contraction:* This augmentation technique helps the network deal with typical variations in inter-beat intervals. To implement it, the input signal is expanded or contracted time by re-sampling it through a standard anti-aliasing low-pass filter. The expansion range was randomly chosen between 0.8 and 1.25.
- *Additive White Gaussian Noise (AWGN):* This augmentation technique aims to make the network robust to standard wireless noise. To implement it, AWGN is added to the signal, with a mean of 0 and a variance randomly chosen between 0.1 and 0.4 of the signal variance.
- *White Gaussian Noise Replacement:* This augmentation technique makes the network robust to erasures in the sensed wireless signal. In the context of WiStress, such erasures may arise from sudden and large body motions like standing up. To implement this technique, we select and replace a random interval inside the signal with white noise. Moreover, to represent large movements, the variance of the noise for this augmentation method was chosen to be 5 to 10 times larger than the signal noise. [11] Unlike other augmentations, this augmentation was applied with a probability of 0.5.
- *Applying random Polynomials:* The final augmentation method aims to represent unseen variations in the heartbeat morphology. To implement this augmentation, we applied a non-linear, random polynomial to all signal samples. Fig. 9 (a) shows a typical wireless signal containing heartbeats, while (b), (c) and (d) are copies of (a) with random polynomials applied. Each color in the figure shows the same pattern in different versions of the signal. As can be seen, the original five patterns in (a) have preserved their shape in the

[9]The number of epochs is a hyper-parameter that is fine-tuned based on the validation dataset.
[10]Note that whenever the input signal is modified via a certain augmentation technique, the corresponding ground-truth IBI of that signal is updated accordingly.
[11]We note that here, unlike the additive version, the ground-truth values in the interval are removed.

other three copies. This shows that while polynomials are a general family of functions,[12] they preserve similar patterns in the signal. Since these patterns are preserved, we can train the network by augmenting signals such as the one in Fig. 9 (a) to those in (b), (c), and (d) without modifying the ground-truth.

5 EVALUATION

We evaluated WiStress's performance for both long-term and short-term stress monitoring. We also evaluated its accuracy and robustness to noise when extracting HRV, and compared it against the baselines described in §4. This section details our performance results.

Participants. We recruited 22 different subjects (6 females), aged 21 to 35, and performed the evaluation in accordance with our IRB protocol. Participants wore their daily attire such as T-shirts, blouses, and button-downs with different fabric materials. We performed our experiments in open offices, study rooms, and living rooms with standard furniture and in co-existence with other wireless technologies (WiFi, LTE, Bluetooth, etc.). We didn't limit the number of people in the room but the subject was always the one closest to the device.

5.1 Experiment 1: Long-Term Stress Monitoring

Experiment. Our first experiment aimed at evaluating WiStress's ability to monitor workload-induced stress over an extended duration. This experiment was performed with 15 out of our 22 participants. During the data collection process, subjects were asked to go about their daily routine (e.g., reading, writing, typing, listening to music, eating) without any constraints. The millimeter-wave sensors were placed around the subjects (e.g., display, desk) within 8 ft of the subject's working area as shown in Fig. 10. We collected 6-hour-long recordings

[12] In fact, polynomials can be arbitrarily close to any continuous function operating on a closed interval [35].

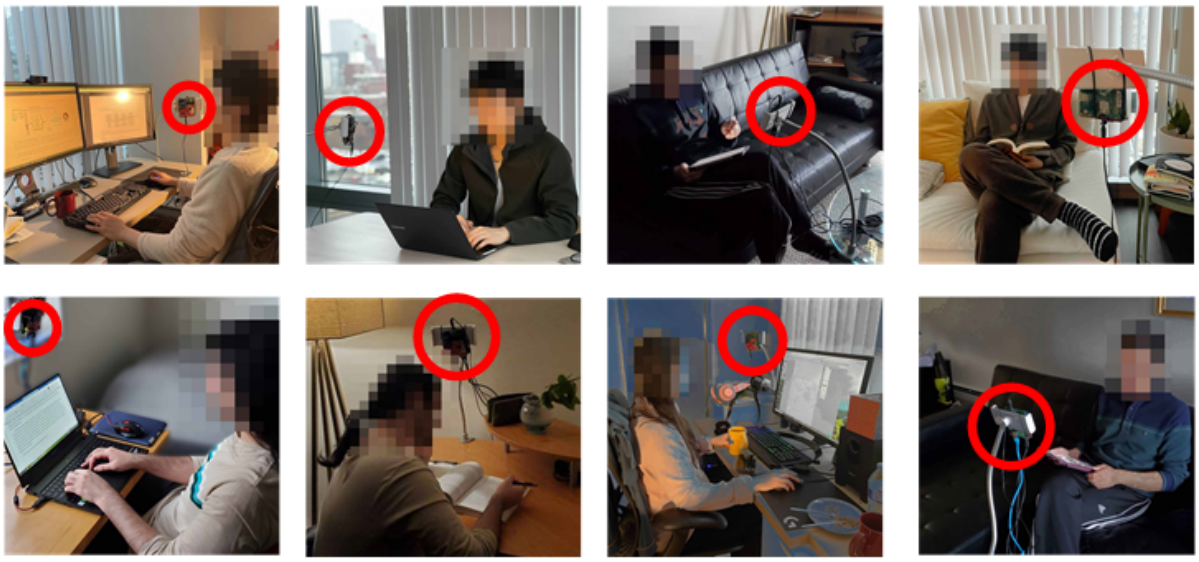
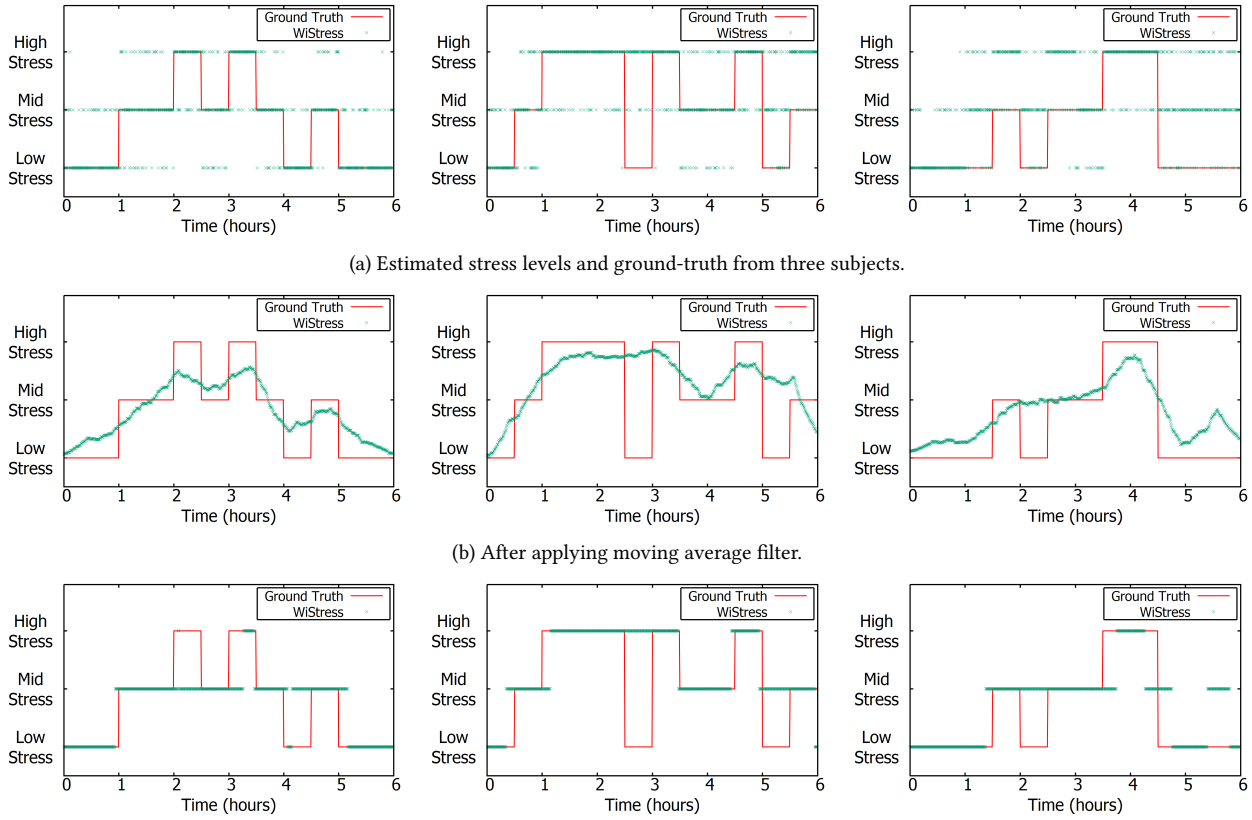


Fig. 10. WiStress setup for long-term stress monitoring. The figure shows the long-term stress monitoring environments. WiStress collected data from 15 subjects for the experiments. WiStress (in red circle) was located on the desk or table for the measurements as shown in the figure.



(a) Estimated stress levels and ground-truth from three subjects.

(b) After applying moving average filter.

(c) After rounding the averaged output.

Fig. 11. Classification Processing Flow. The figure shows the ground-truth stress level (red) and WiStress's estimated stress level (green) for three different subjects across 6 hour experiments.

for each subject, during which the subjects were free to leave and return to the area near the WiStress sensor. Each of the subject was performing a variety of tasks due to external work-related events (e.g., paper deadline, exam night study session, meeting preparation).

Ground Truth: To obtain ground-truth measurements of the subjects' workload stress levels, we asked each subject to complete the NASA-TLX (Task Load Index) survey once every 30 minutes during the experimental period. The NASA-TLX [57] survey is an assessment tool widely used to quantify a subject's workload stress level [57, 100, 131]. It consists of six rating categories: Mental Demand, Physical Demand, Temporal Demand, Overall Performance, Effort, Frustration Level. Each category is rated on a 100-point scale with 5-point steps and the total rating is summed up (600 points) and used as a stress level indicator [28, 56, 61]. Similar to past literature, we used this survey to classify between two and three stress levels. For the two-class training, we classified between lower (1-300 points) and higher (301-600 points) stress levels. For three-class training, we classified into three levels: low (1-200 points), mid (201-400 points), high (401-600 points).

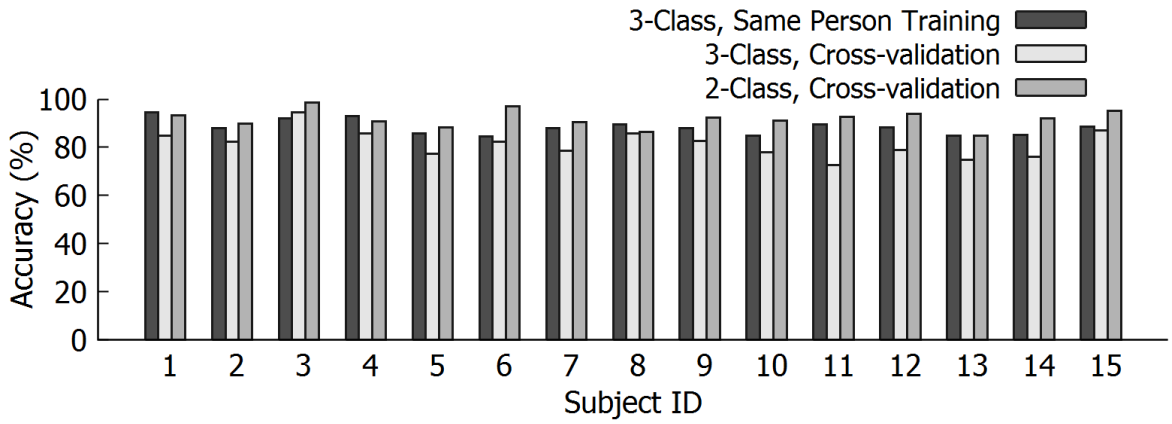


Fig. 12. Classification Accuracy in Long-term Stress Monitoring. The figure plots WiStress's classification accuracy in long-term stress experiments for 15 human subjects. Each cluster represents a single subject and the corresponding metrics.

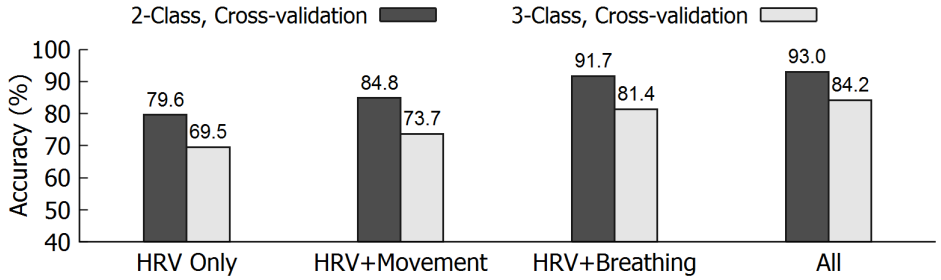


Fig. 13. Impact of Different Features on Accuracy. The figure plots WiStress's classification accuracy when using different features. While HRV is the main contributor in the classifier's performance, both movement and breathing features contribute meaningfully to WiStress's overall classification accuracy.

Results. We evaluated WiStress's ability to estimate stress levels in long-term measurements in two different regimes: the first involves training and testing for the same subject, while the second regime involves testing on one subject while training on all other subjects. In the first regime, the dataset from each subject was divided into training and testing sets. For the second regime, the model was trained on 14 other subjects and evaluated on the subject it hasn't been trained on (leave-one-out cross-validation). It is important to note that the measurements were conducted without any constraints on the subjects' activities. Across the experimental trials, WiStress remotely captured the reflections from the subject and automatically extracted heart rate variability and respiration features.

We computed two evaluation metrics, namely, two-, and three-class classification accuracy, similar to prior work on long-term stress monitoring [32, 50, 51, 120]. Fig. 11 (a) plots the estimated stress levels and ground-truth from three subjects. As expected, WiStress's estimation fluctuates during the measurements. We note that, since the ground truth is sampled only once every 30 minutes, it is not capable of fully tracking the variation of stress level within the 30-minute duration. To remedy that, we applied a moving average filter over WiStress's raw prediction to achieve its representative levels every 3 minutes of WiStress's estimation in Fig. 11 (b). The final estimation was determined by rounding to the closest of the three levels the output as in Fig. 11 (c).

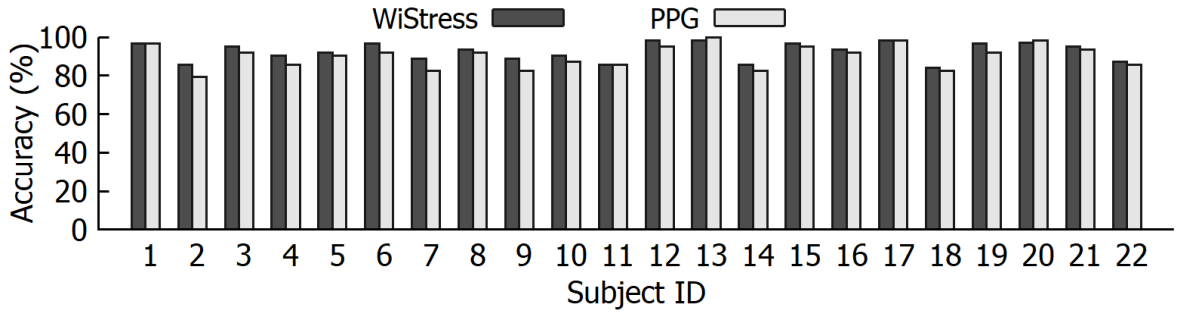


Fig. 14. Classification Accuracy in Short-term Stress Monitoring. The figure plots WiStress's classification accuracy and the PPG-based classification accuracy in short-term stress experiments for 22 human subjects.

Fig. 12 plots the metrics across all 22 subjects. Each cluster represents a single subject and the corresponding metrics. We make the following remarks:

(1) When trained and tested on the same subject, WiStress achieves high average accuracy (>88%) for the 3-class case across all subjects. Three subjects achieve 90+% accuracy.
(2) When tested on a new subject, the accuracy for the 3-class case drops but it remains around or above 80%. It is also worth noting that WiStress has similar or higher accuracy than the prior work using contact-based sensors (<80%) [32, 50, 51, 120] for real-life stress measurements. This shows that WiStress's performance is comparable to contact-based systems.
(3) When trained and tested for 2-class classification, WiStress achieves 91.7% average accuracy. Additionally, WiStress's accuracy remains above 90% for eleven out of fifteen subjects. This result demonstrates WiStress's ability to estimate stress levels without any constraints on the subjects.

Breakdown of WiStress's Overall Accuracy. To better understand the importance of each stress feature described in §3.3 (HRV, breathing, movement), we study their impact on WiStress's overall stress classification accuracy. To do this, we trained and tested four variants of WiStress's classifier with a subset of these features: 1) using only HRV features; 2) using HRV and moving features; 3) using HRV and breathing features; 4) using all features. We used the same data collected from the 15 subjects and trained our classifiers on the 2-class and 3-class cross-validation scenarios.

Fig. 13 shows the accuracy for each of these implementations. We make the following remarks:

- The results show that a classifier using only HRV achieves an accuracy of 79.6% and 69.5% in the 2-class and 3-class scenarios respectively.
- When combined with HRV, each of the movement and breathing features meaningfully contributes to improving the classification performance. For example, in the 3-class experiment, using movement and breathing features increases the classification accuracy by 5.2% and 11.1% respectively compared to HRV alone.
- Finally, the combination of HRV, movement, and breathing yields the highest classification accuracy for both the 2-class and 3-class scenarios. This demonstrates that WiStress's stress classification module benefits from all three types of features.

5.2 Experiment 2: Short-term Stress Monitoring

Experiment. Our short-term stress monitoring experiment was performed with all 22 participants we recruited. Unlike the long-term experiment which was performed in uncontrolled environments, here, we evaluated WiStress's performance with pre-designed stress-inducing tasks, which provide more objective ground-truth in a controlled environment similar to [11, 12, 103, 118]. During the data collection process, subjects were asked to sit in a chair within 4 ft from the sensor and remained quasi-static during the experimental trials. The millimeter-wave sensor was placed in front of the subjects and aimed at the sternum. To collect measurements across three different stress levels, we have induced three levels of stress and measured the individual responses with WiStress. Three sessions with low, medium, and high stress were defined; each session consisted of a baseline, stress, and recovery stages. Each session took 15 minutes. The baseline and recovery stages were the same for the three sessions: the subjects meditated while listening to relaxing music. For the three different stress sessions, we used meditation, 1-Back, and 2-Back[13] Tasks [104] to induce low, medium, and high stress. To compare WiStress's performance to contact-based systems, a PPG device was used in parallel as described in §4.3. Training and testing were done on completely different subjects, and testing was performed using leave-one-out validation. We compared WiStress's stress classification accuracy against the PPG baseline.

Results. Fig. 14 plots the 3-class stress classification accuracy across all 22 subjects for the short-term stress measurements. In quasi-static environments, the accuracy of WiStress and PPG remains around or above 90% on average. Note that WiStress's 3-class classification accuracy is higher here than in the long-term stress monitoring experiment. This is expected because the short-term experiment was performed in a more controlled manner. This improved accuracy is also in line with prior literature that compared the accuracy of stress monitoring in controlled and uncontrolled environments [87].

Another interesting observation is that WiStress has similar or even higher accuracy than PPG across different subjects. Specifically, WiStress can achieves average accuracy of 92.6%, which is 2.5% higher than PPG's 90.1%. This is likely because WiStress also uses respiration signals which cannot be extracted by the PPG-based baseline. To better understand WiStress's performance against PPG, we compared the two methods with and without breathing features. In particular, we implemented a second baseline that combines PPG with respiration signals obtained from WiStress; we also implemented a variant of WiStress that relies entirely on HRV measurements for stress classification. The results are shown in table 2.

Table 2. Average Classification Accuracy of WiStress and PPG. The table compares the accuracy of WiStress and the PPG-based modality with or without the breathing features.

	w/o Breathing Features	w/ Breathing Features
WiStress	82.9	92.6
PPG	90.1	94.3

As the table demonstrates, removing breathing features from WiStress results in 82.9% accuracy, which is 7.2% lower than PPG. This is because wireless signals naturally have lower signal-to-noise ratio (SNR) than PPG signals, resulting in less accurate HRV measurements, thereby lower stress classification accuracy. However, by including the breathing features, WiStress's performance jumps to 92.6%, which is 2.5% higher than PPG without breathing features, and a mere 1.7% below PPG with breathing features.[14] These experiments show that while the wireless modality typically has a lower SNR to extract the HRV, we can make up for it by further extracting the breathing features extracted from the signal.

[13]Whenever the subject's answer was incorrect, annoying sounds (high-pitched scream, siren) were played.

[14]Since PPG itself does not provide breathing features, for the "PPG with breathing features", experiment we added the breathing features extracted from our wireless sensors.

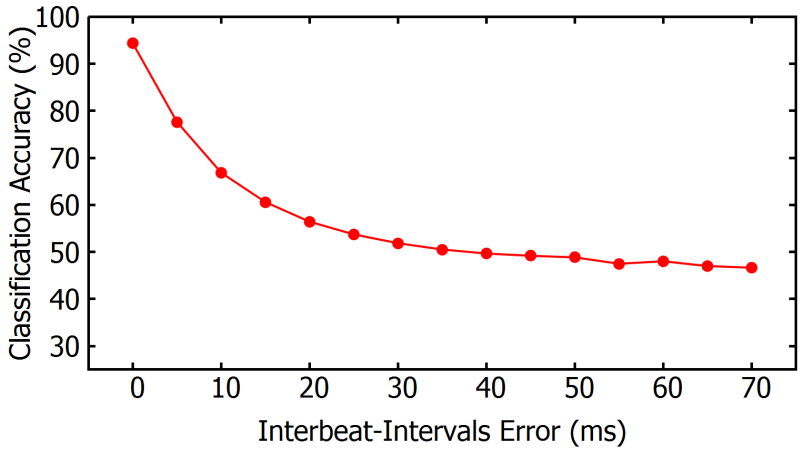


Fig. 15. Classification Accuracy vs. IBI error. The figure plots the classification accuracy between three classes as a function of the median IBI error. As the IBI error increases, the classification accuracy drops. We note that the existence of other features (breathing and movements) keeps the accuracy above random guess (33%).

Two additional points are worth noting:

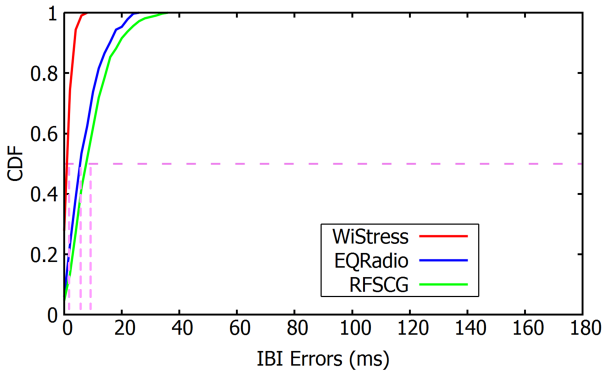
- In the short-term classification experiment, we do not leverage WiStress's movement-extraction features. This is because in this experiment, subjects are asked to remain quasi-static as described earlier.
- The accuracy of WiStress with and without breathing features is higher in the short-term experiment than it is in the long-term stress experiment. As described earlier, this is due to the more controlled nature of this experiment and is in line with prior literature.

5.3 Accuracy of IBI Extraction

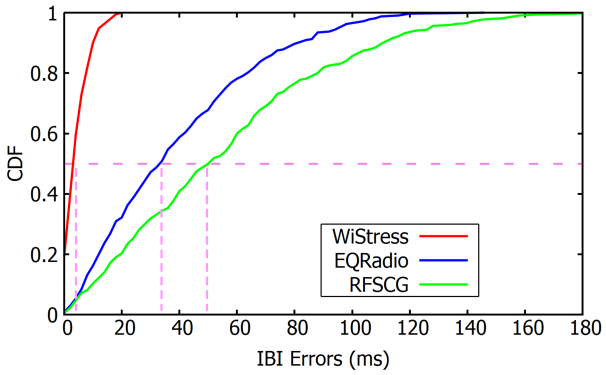
In this section, we evaluate the importance of accurately extracting the inter-beat interval (IBI) on WiStress's overall stress classification accuracy, and compare its IBI extraction algorithm to start-of-the-art baselines.

5.3.1 Impact of IBI Error on Stress Classification Accuracy. We are interested in understanding how much the IBI error impacts the WiStress's final performance. To do so, we performed controlled experiments to emulate the IBI error by adding different levels of random noise, and assessed the impact of such error on the overall classification error. We took the ground truth heartbeats derived from our PPG sensor and added different levels of White Gaussian noise with zero mean and standard deviation varying from 0 and 90 ms, similar to past work [128] to make noisy IBIs. We then ran our person-independent stress classification module on one subject from the short-term experiment using these noisy IBI values. For this experiment, we did not add noise to the other features (breathing and movement), and they remained intact.

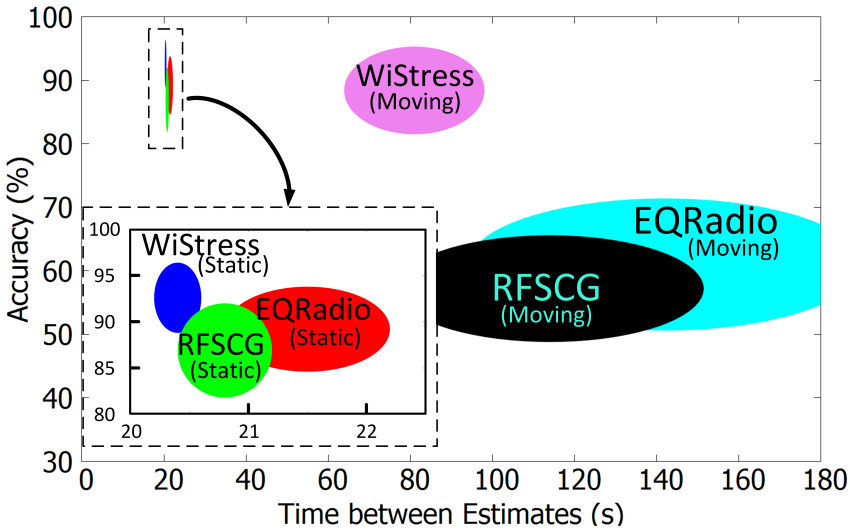
Fig. 15 plots the the stress classification accuracy against average IBI error. Each data point shown in the figure was averaged over 100 random experiments. As can be seen, to maintain an accuracy of 70% or more, the IBI error needs to be below 10 ms. As the IBI error approaches 30 ms and higher, the classification accuracy drops to around 50%. This shows that in order to maintain reasonable stress classification performance, we need to achieve IBI accuracy of below 10 milliseconds. Finally, it is worth noting that since WiStress uses two groups



(a) CDF of IBI error when the subject is static



(b) CDF of IBI error when the subject is moving



(c) Stress Classification Accuracy vs. Time between Estimates. Each method is represented using an oval centered around its median (time, accuracy) and whose axes represent the standard deviation along the respective dimension.

Fig. 16. Classification Accuracy and IBI error. The figure compares WiStress to different baselines in terms of IBI errror, stress classification accuracy, and time between estimates.

of other features, namely breathing and movement features, even with very high IBI error, the classification accuracy remains well above the 33% random guess.

5.3.2 Comparison with State-of-the-Art Baselines. Next, we would like to quantify the benefits of WiStress's IBI extraction network (WiStress-Net) against two state-of-the-art baselines in RF sensing (RFSCG [52], EQRadio [128]) as described in §4.3. To do so, we evaluated the IBI extraction pipeline of each system when the user is static and moving. We conducted two (static, moving), 20-minute experiments, where a subject was sitting 3 ft away from the millimeter-wave device. During the static experiment, the user was asked to remain quasi-static (slowly moving hands or head was allowed), and during the moving experiment, the user was free to move in place (unconstrained shaking their head, moving their hands, speaking, and drinking were allowed). The ground truth IBI's were collected by attaching the PPG sensor to the subject's big toe. To ensure an accurate ground-truth, we asked the subject not to perform any abrupt or shaking motions with that foot throughout the experiment.

Fig. 16 (a) and (b) plot the empirical CDF of WiStress's IBI error against the baselines for the static and moving scenarios, respectively. We make the following remarks:

- Overall, WiStress maintains a low median error (<4ms for static, <6ms for moving) for both experiments.
- WiStress shows a modest improvement over the baselines when the user is static (EQRadio:8ms, RF-SCG:10ms). To see why, recall from Fig. 3 that the heartbeat pattern may change during the same experiment. However, both RFSCG and EQRadio rely on learning a single template to match heartbeats throughout the experiment, which means they cannot adapt to the heartbeat pattern changes. On the other hand, WiStress detects heartbeats by relying on *local* self-similarities, yielding more accurate IBI estimates.
- When the user is allowed to move, Fig. 16(b) shows that WiStress outperforms the two baselines by a significant margin (EQRadio:36ms, RFSCG:48ms). This is because WiStress is able to correctly reject motion-based artifacts from the measurement signal, and only focus on segments that contain heartbeat signals without motion artifacts. In contrast, the baselines mostly rely on the user being static and provide little to effectively combat the motion artifacts.
- It is also worth noting that the accuracy of WiStress in moving experiment exceeds the accuracy of the baselines in static experiments, which demonstrates that WiStress is highly robust to real-life artifacts.

Finally, one might wonder whether WiStress's improved IBI extraction accuracy is the result of a more conservative motion rejection mechanism than the baselines. In the next section, we demonstrate that this is not true and that WiStress is successful at obtaining both more frequent and more accurate stress estimates than the baselines.

Implications on Stress Monitoring. So far, we have compared WiStress to the baselines based on the IBI extraction accuracy. Next, we would like to compare the three methods based on (a) the accuracy of stress monitoring and (b) frequency of making stress estimates. Note that since the baselines cannot perform stress monitoring, we augment them with the rest of WiStress's pipeline (including feature extraction from IBI, breathing, and motion) for fair comparison.

We repeated the same experiment as that of the short-term stress monitoring assessment (described §5.2). However, rather than asking subjects to remain static throughout the experimental trials, we allowed them to move in a manner similar to that described above. We performed 12 experimental trials in total with four subjects, each lasting for four minutes.

Fig. 16(c) shows the 3-class stress classification accuracy as a function of the time between stress estimates for WiStress, RF-SCG, and EQ-Radio. For each of the three methods, we the results are shown separately for the static and moving scenarios described above. For each method and scenario, the result is represented as an oval centered around the median (time, accuracy) and the two axes of the oval represent the corresponding standard deviation (i.e., along time and accuracy). It is worth noting that each system has their own motion rejection modules, which impacts the time between measurements on the x-axis.[15] We make the following observations:

(1) *WiStress vs baselines (static).* When static, all three methods are able to make a measurement once every 20-22 seconds (i.e., the duration of the sliding window), with 85-95% average accuracy.[16] This means that if a subject is static, all three IBI extraction algorithms methods can achieve good accuracy (>85%) for stress monitoring (once the algorithms are augmented with WiStress's stress classification architecture described in §3.3). However, we note that continuous stress monitoring of a static user is not realistic, as it is unlikely that a person is static for longer than a few minutes at a time in an everyday environment [65].

[15]Since EQRadio does not describe its motion rejection method, we assumed the authors leveraged the method described in their earlier work [10] for motion rejection.

[16]WiStress has a modest improvement over the baselines both in terms of accuracy and frequency of the assessments, thanks to a better HRV extraction method, as shown in Fig. 16(a).

(2) *WiStress vs baselines (moving).* When the user is free to move, all three methods require more time between stress estimates, with WiStress outputting an estimate every 70-90 seconds, with EQ-Radio and RF-SCG trailing behind at around 100-180 seconds. While these numbers are lower compared to the static case, they are still well within the range of acceptable frequencies for the purpose of stress monitoring [102]. The real difference between the methods is in their accuracy, with WiStress maintaining its 85-95% accuracy level, and RFSCG and EQRadio dropping to 50-70%. This accuracy is likely due to the breathing and motion features. This comparison shows that in real, everyday environments where the user is free to move, WiStress will significantly outperform the baselines, and it is the only method among the three that enables accurate, continuous stress monitoring. We note that this result is in agreement with our previous analysis on Fig. 16 (b). This demonstrates that WiStress's robustness of IBI extraction translates to a higher performance in stress monitoring over the baselines.

(3) *WiStress static vs. WiStress moving.* Since user motion artifacts in the moving experiments can overwhelm the subtle heartbeat and respiration signals, WiStress strategically rejects these portions as described in §3.2. This rejection has two implications; First, it introduces a certain degree of sparsity in the data samples, which ultimately translate to less frequent stress measurements. Second, it helps keep the original accuracy of the measurements as in the static case by removing the low-quality portions which convey no useful information. Both implications are reflected in Fig. 16 (c), where we can see an increase in the latency of WiStress's stress estimation from every 20 seconds in the static case to every 80 seconds in the moving case, while the accuracy is maintained at around 85-95%.

5.4 Robustness and Generalizability

5.4.1 Robustness to Orientation and Distance. We evaluated WiStress's IBI accuracy as a function of orientation and distance. First, we performed an the experiment where we varied the distance between WiStress and the subject of interest from 2-12 ft. Then, we repeated the IBI extraction for four different orientations: facing the device, left, right, and back. We performed each experimental trial three times for each distance-orientation pair.

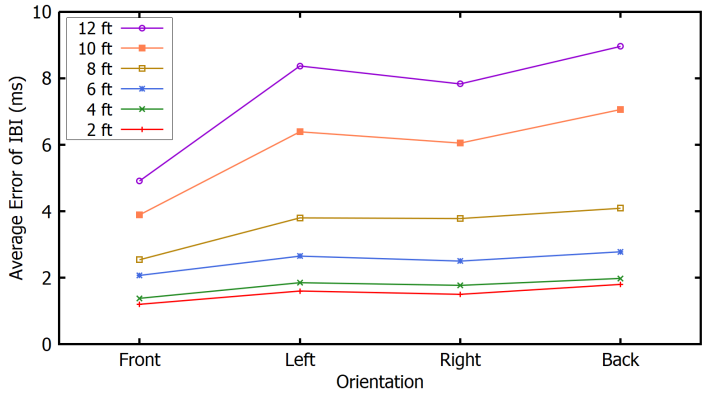
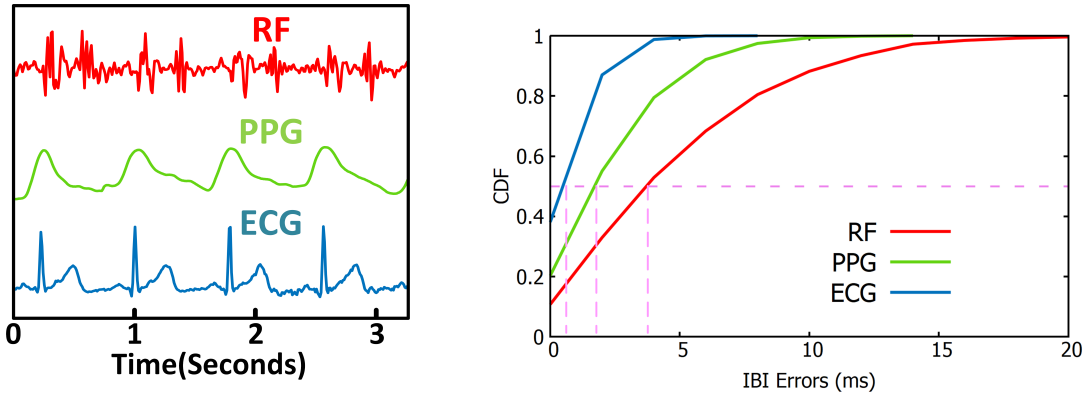


Fig. 17. Error in IBI vs. orientation and distance. The figure shows the average error of WiStress's IBI estimation for different orientations (front, left, right, back) and distances (2-12 ft).

Fig. 17 plots the average error of WiStress's IBI estimation across distance and orientation. It shows that the average error remains below 4 ms if the subject is within 8 ft distance. When the subject faces back, it shows the



(a) RF, PPG, and ECG measurements for generalizability test

(b) CDF of IBI error for different modalities

Fig. 18. WiStress-Net's Generalizability. (a) The figure plots the waveforms of three modalities. Three different signals have different shapes with the same period. WiStress-Net has not been trained on PPG and ECG for the test. (b) The figure plots the CDF of the IBI errors for WiStress-net for each of the three modalities.

highest error, similar to prior work [10]. Yet, even at 12 feet, the average error remains less than 10 ms for all orientations.

5.4.2 Generalizability of WiStress's HRV Extraction. Next, we evaluated the generalizability WiStress's IBI extraction module (WiStress-Net, §3.2.3). Specifically, we ran experiments to test whether our network can accurately extract IBIs from heartbeat signal modalities other than RF. To this end, we evaluated our IBI extraction model on modalities other than those our network had been trained on. Specifically, without any additional training, we applied our network to ECG and PPG signals, which are 1D heartbeat signals that have periodicity but are measured through different physiological mechanisms. We measured the heartbeat signals using RF, ECG, and PPG modalities as shown in Fig.18(a). Across these experimental trials, the subject was sitting about 4ft away from WiStress's antenna and remained quasi-static during the measurements. The ECG and PPG signals were also measured from the subject's chest and finger, using hardware described in §4.1. We conducted 5 experimental trials, each lasting for 10 minutes. We compared the manually labeled ground-truth to the extracted IBIs from our learning model.

Fig.18 (b) plots the CDF of the IBI error for RF, PPG, and ECG signals. Each of the empirical CDFs combine all the data across all 5 trials. We make the following remarks:

(1) All the modalities achieve low 90^{th} percentile errors ($\leq$10ms), which corresponds to only five sample points. This demonstrates that WiStress's IBI extraction model can indeed learn temporally local self-similarities regardless of the signal modality and that it extends to other heartbeat signal modalities that exhibit periodicity.
(2) Interestingly, the ECG and PPG modalities show smaller median errors ($\leq$ 2ms) compared to the RF modality (4 ms). This is expected since contact-based modalities typically have a higher SNR compared to RF.

These results demonstrate that WiStress-Net not only enables wireless IBI extraction, but is general enough to be applied for other heartbeat sensing modalities.

5.4.3 Robustness to Different Daily Activities. In this experiment, we aimed to understand the potential of monitoring stress during a user's everyday activities. In particular, recall that WiStress makes a best effort

to continuously extract IBIS and that it discards measurements with significant motion contamination. Hence, we are interested in evaluating the frequency of estimating stress across various daily activities. To do this, we measured millimeter-wave reflections while a subject performed various daily activities (sleeping, reading, speaking, typing, workout) for 40 minutes each. The first 10-minute data were discarded away because the subject was more aware of the measurement in the early phase.

Each measurement setup is as follows:

(1) **Sleeping**. The subject was lying down in bed and under a blanket. The device was installed on the headboard of the bed and measured the signal 4 ft away from the subject.
(2) **Reading**. The subject was reading a magazine on a sofa. The average speed of turning page was 53 seconds/page. The measurement distance was 5 ft.
(3) **Speaking**. The subject had a virtual meeting with the colleagues and led the meeting. The subject used 42 words/min in average. The measurement distance was 3 ft.
(4) **Typing**. The subject typed a document while sitting on a chair. The average typing speed was 51 words/min. The measurement distance was 3 ft.
(5) **Workout**. The subject worked out alone. The routine was composed of squat, lunge, push up, and jumping jack. The break time didn't exceed 5 minutes in total. The measurement distance was 6-8 ft.

For each of the experimental trials, we computed the duration between consecutive successful stress estimates. Then, for each of activity, we computed the mean and 90^{th} percentile of those durations, and show the results in Table 3.

Table 3. Average latency between estimates during daily activities. The figure shows how often we can estimate the stress level across various daily activities.

Task	Minutes between two estimates	
	Mean (min)	90^{th} Percentile (min)
Sleeping	0.34	0.60
Reading	1.19	2.63
Speaking	2.56	4.02
Typing	3.12	5.13
Workout	N/A	N/A

We make the following points:

(1) Except for the workout activity, for all other activities, WiStress can extract infer the person's stress level at least once every 3.12 minutes on average. Since a typical stress episode lasts on average for around 11.5 minutes [102], this indicates that WiStress can capture standard stress episodes across daily activities (and even output multiple stress estimates for each episode).
(2) As expected, during low-motion activities (such as sleeping or reading), the number of opportunities for extracting stress levels is higher, resulting in shorter mean and 90^{th} percentile durations between estimates.
(3) The 90-th percentiles show that the tail of the time between measurements is acceptable. In particular, except for workout, the 90-th percentile of time between two measurements is at most 5.13 minutes.

Impact of Subtle Movements. Our final result aims to provide insights into the robustness of WiStress's heartbeat extraction to small movements, like moving head or hand slowly. To this end, we ran experiments where we compare the SSM output in four different scenarios with different levels/types of movements.

Fig. 19 plots the SSM for each of the four scenarios, as well as the corresponding wireless reflection (that was fed into WiStress-Net), shown as the black curve. The figure also depicts the detected heartbeats using

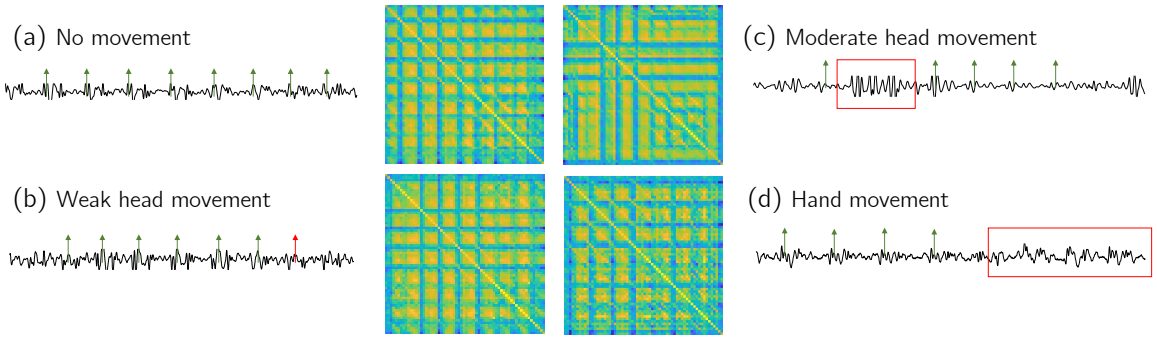


Fig. 19. Effect of small movements on the input, SSM, and detected heartbeats. The inputs are the wireless reflection data shown as black curves. The correctly detected heartbeat outputs are shown in green arrows. The red arrow indicates a missed heartbeat. The rectangles indicate the discarded regions.

upward facing arrows. Green arrows indicate detected heartbeats that match with manual annotation, and red arrows indicate missed heartbeats. In (a), the subject is not moving. In (b), the subject is slowly moving their head horizontally about 90 degrees over a 15 second period, the middle of which is shown in the figure. In (c), the subject moves their head vertically from looking ahead to looking down in about 2 seconds. The movement period is shown in a red box. In (d), the subject is moving their right hand from the chair armrest to a resting table on the table. Similarly, the movement period shown in a red box.

We can see that when there is no movement or the movement is weak (a-b), WiStress performs very well, missing only one annotated heartbeat in (b). This is also reflected in the structure of the SSM, were we can see similar structures repeating, indicating the presence of heartbeat. While moderate movements (c-d) introduce some missingness in the data, WiStress -Net still correctly identifies and rejects these regions, while still correctly identifying the annotated heartbeats. This is also reflected in the SSMs, where the patterns disappear around the areas corresponding to motion artifacts. These results demonstrate that WiStress is robust to small movements and can automatically learn and discard segments where the movement causes significant motion contamination.

These micro-benchmarks also demonstrate that WiStress is successful in learning how to balance between robustness and availability. As described in §3.2 and §3.3, WiStress keeps or discards certain regions of the wireless data before feeding them into its stress classification module. Specifically, data samples that pass a certain quality threshold are kept, and the rest are discarded. The value of this threshold poses a quality-quantity trade-off: a lower threshold leads to more, possibly erroneous data samples, while a higher threshold would result in less, yet more accurate samples. In principle, it is possible to further adapt this threshold based on the end use-cases.

6 LIMITATIONS & EXTENSIONS

In this section we describe the limitations of WiStress and avenues for extending it:

(1) *Range.* Our current evaluation of WiStress was performed up to a distance of around 4 meters. In the future, it is desirable to extend the operation range by incorporating techniques such as beamforming with more antennas or using more sensitive hardware.
(2) *High-Energy Activities.* While WiStress can continuously monitor a subject's stress in everyday scenarios like typing, talking, or reading as well as deal with intermittent large motions, it cannot extract stress measurements during continuous, high-energy motion scenarios like workout sessions (since it needs to

discard time segments with high variance). It would be interesting to explore extensions of WiStress that can operate even while the user is performing high-energy activities.

(3) *Number of Users.* Our current implementation of WiStress works for one user at a time. In principle, it is possible to extend WiStress to accommodate multiple subjects at the same time through digital beam forming as shown by past work [10, 130].

(4) *Longitudinal Studies.* In this paper, we demonstrated WiStress's ability to monitor stress for up to 6-hour-long experiments. It would be interesting to use WiStress to monitor stress over longer periods, like multiple days, weeks, or even months.

Finally, it is worth noting that this paper explored WiStress's ability to monitor both naturally occurring stress (in the long-term studies) as well as induced stress (in the short-term studied) by leveraging prior stress models and classifiers. It would be interesting to adapt and evaluate WiStress with other stress models from the literature [87] by considering different stress classifiers as well as by different stress models (i.e., beyond the NASA TLX or the Back tests). Moreover, our implementation of WiStress has selected training hyper-parameters that balance between availability and robustness (as detailed in §5.4.3), and it is worth exploring how the trade-off between these metrics might be adapted to various use cases. Finally, it would be interesting to explore the benefits and use cases of multi-modal extensions of WiStress, where it is combined with other sensor data (e.g., accelerometers, voice) in stress monitoring applications.

7 CONCLUSION & FUTURE DIRECTIONS

In this paper, we presented WiStress, the first contactless stress monitoring system that can infer user stress levels from wireless signals. WiStress works correctly without requiring any contact with the user's body, or the user being aware of the system. By automatically and opportunistically extracting vital signs and other stress-related features from a nearby user and designing a novel pipeline for extracting stress-related features from wireless signals, WiStress enables long-term monitoring of user stress levels.

In comparison to existing approaches for stress monitoring – which either solicit user input or require users to wear on-body sensors (e.g., ECG, GSR, PPG) – WiStress offers a more seamless, transparent, and convenient modality for long-term stress monitoring. It opens up opportunities for monitoring stress in workplaces [72] or academic environments [109, 110], where the sensor can be placed in a room or on a desk to monitor nearby users. Such deployments can help boost productivity and performance and reduce burnout [44, 53, 88, 99, 106], as well as inform intervention mechanisms to support student and worker mental health and well being [34]. WiStress can also be incorporated into smart devices (e.g., screens, kiosks, TVs, smart home assistants) to help understand and respond to user stress levels. For example, it can be used in studies for user experience evaluation [19, 115] or in everyday environments to enable interactive capabilities (adapt their tone, colors, etc.) based on user stress levels [75]. Additionally, WiStress sensors also be used to assess stress during sleep [54] and help improve sleep quality. Beyond smart devices and smart environments, WiStress can also be incorporated in social robots to help them asses user stress, improve interventions for aging [25] and child learning [15], and support long-term psychological well-being.

More generally, WiStress paves the way towards transparent and ubiquitous stress monitoring systems, with applications spanning smart homes, human-computer interaction, and mental health and well-being.

ACKNOWLEDGMENTS

We thank the anonymous reviewers and the Signal Kinetics group for their feedback on the paper. This research is sponsored by an NSF CAREER Award (CNS-1844280), the ONR YIP Award (N00014-19-1-2325 A00001), the Sloan Research Fellowship, NTT DATA, Toppan, Toppan Forms, the MIT Media Lab, and the Abdul Latif Jameel Clinic for Machine Learning in Health (J-Clinic) at MIT.

REFERENCES

[1] [n.d.]. *BioStamp RC User Manual.* https://www.mc10inc.com/hubfs/UserManual.pdf

[2] [n.d.]. The SK Learn Library.

[3] [n.d.]. Texas Instruments ADS1299. https://www.ti.com/tool/ADS1299EEGFE-PDK. Accessed: 2021-02-13.

[4] [n.d.]. Texas Instruments AFE4400SPO2EVM. https://www.ti.com/tool/AFE4400SPO2EVM. Accessed: 2021-02-13.

[5] [n.d.]. Texas Instruments iwr1443BOOST. https://www.ti.com/store/ti/en/p/product/?p=IWR1443BOOST. Accessed: 2021-02-13.

[6] [n.d.]. Texas Instruments mmWave Studio. http://www.ti.com/tool/MMWAVE-STUDIO. Accessed: 2021-02-13.

[7] U Rajendra Acharya, K Paul Joseph, Natarajan Kannathal, Choo Min Lim, and Jasjit S Suri. 2006. Heart rate variability: a review. *Medical and biological engineering and computing* 44, 12 (2006), 1031–1051.

[8] Fadel Adib, Chen-Yu Hsu, Hongzi Mao, Dina Katabi, and Frédo Durand. 2015. Capturing the human figure through a wall. *ACM Transactions on Graphics (TOG)* 34, 6 (2015), 1–13.

[9] Fadel Adib, Zach Kabelac, Dina Katabi, and Robert C Miller. 2014. 3d tracking via body radio reflections. In *11th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 14)*. 317–329.

[10] Fadel Adib, Hongzi Mao, Zachary Kabelac, Dina Katabi, and Robert C Miller. 2015. Smart homes that monitor breathing and heart rate. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 837–846.

[11] Jonathan Aigrain, Séverine Dubuisson, Marcin Detyniecki, and Mohamed Chetouani. 2015. Person-specific behavioural features for automatic stress detection. In *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, Vol. 3. IEEE, 1–6.

[12] Jonathan Aigrain, Michel Spodenkiewicz, Severine Dubuisson, Marcin Detyniecki, David Cohen, and Mohamed Chetouani. 2016. Multimodal stress detection from multiple assessments. *IEEE Transactions on Affective Computing* 9, 4 (2016), 491–506.

[13] Mustafa Al'Absi, Stephan Bongard, Tony Buchanan, Gwendolyn A Pincomb, Julio Licinio, and William R Lovallo. 1997. Cardiovascular and neuroendocrine adjustment to public speaking and mental arithmetic stressors. *Psychophysiology* 34, 3 (1997), 266–275.

[14] Ane Alberdi, Asier Aztiria, and Adrian Basarab. 2016. Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review. *Journal of biomedical informatics* 59 (2016), 49–75.

[15] Safinah Ali, Tyler Moroso, and Cynthia Breazeal. 2019. Can children learn creativity from a social robot? In *Proceedings of the 2019 on Creativity and Cognition*. 359–368.

[16] Urs Anliker, Jamie A Ward, Paul Lukowicz, Gerhard Troster, Francois Dolveck, Michel Baer, Fatou Keita, Eran B Schenker, Fabrizio Catarsi, Luca Coluccini, et al. 2004. AMON: a wearable multiparameter medical monitoring and alert system. *IEEE Transactions on information technology in Biomedicine* 8, 4 (2004), 415–427.

[17] Christoph Hoog Antink, Simon Lyra, Michael Paul, Xinchi Yu, and Steffen Leonhardt. 2019. A broader look: Camera-based vital sign estimation across the spectrum. *Yearbook of medical informatics* 28, 1 (2019), 102.

[18] Muhammad Arsalan, Avik Santra, and Christoph Will. 2020. Improved contactless heartbeat estimation in FMCW radar via Kalman filter tracking. *IEEE Sensors Letters* 4, 5 (2020), 1–4.

[19] Javier A Bargas-Avila and Kasper Hornbæk. 2011. Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 2689–2698.

[20] Gerald Bauer and Paul Lukowicz. 2012. Can smartphones detect stress-related changes in the behaviour of individuals?. In *2012 IEEE international conference on pervasive computing and communications workshops*. IEEE, 423–426.

[21] Shahina Begum, Mobyen Uddin Ahmed, Peter Funk, Ning Xiong, and Bo Von Schéele. 2006. Using calibration and fuzzification of cases for improved diagnosis and treatment of stress. In *8th European Workshop on Case-based Reasoning in the Health Sciences, Turkey 2006*. 113–122.

[22] Fabrizio Benedetti, Michele Lanotte, Leonardo Lopiano, and Luana Colloca. 2007. When words are painful: unraveling the mechanisms of the nocebo effect. *Neuroscience* 147, 2 (2007), 260–271.

[23] Andrey Bogomolov, Bruno Lepri, Michela Ferron, Fabio Pianesi, and Alex Pentland. 2014. Daily stress recognition from mobile phone data, weather conditions and individual traits. In *Proceedings of the 22nd ACM international conference on Multimedia*. 477–486.

[24] Frédéric Bousefsaf, Choubeila Maaoui, and Alain Pruski. 2013. Remote assessment of the heart rate variability to detect mental stress. In *2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops*. IEEE, 348–351.

[25] Cynthia L Breazeal, Anastasia K Ostrowski, Nikhita Singh, and Hae Won Park. 2019. Designing social robots for older adults. *Natl. Acad. Eng. Bridge* 49 (2019), 22–31.

[26] Ronald J Burke and Esther Greenglass. 1993. Work stress, role conflict, social support, and psychological burnout among teachers. *Psychological reports* 73, 2 (1993), 371–380.

[27] Yekta Said Can, Niaz Chalabianloo, Deniz Ekiz, and Cem Ersoy. 2019. Continuous stress detection using wearable sensors in real life: Algorithmic programming contest case study. *Sensors* 19, 8 (2019), 1849.

[28] Alex Cao, Keshav K Chintamani, Abhilash K Pandya, and R Darin Ellis. 2009. NASA TLX: Software for assessing subjective mental workload. *Behavior research methods* 41, 1 (2009), 113–117.

[29] Tarani Chandola, Eric Brunner, and Michael Marmot. 2006. Chronic stress at work and the metabolic syndrome: prospective study. *Bmj* 332, 7540 (2006), 521–525.
[30] Xinlei Chen, Haoqi Fan, Ross Girshick, and Kaiming He. 2020. Improved baselines with momentum contrastive learning. *arXiv preprint arXiv:2003.04297* (2020).
[31] Komkrit Chooruang and Pongpat Mangkalakeeree. 2016. Wireless heart rate monitoring system using MQTT. *Procedia Computer Science* 86 (2016), 160–163.
[32] Matteo Ciman and Katarzyna Wac. 2016. Individuals' stress assessment using human-smartphone interaction analysis. *IEEE Transactions on Affective Computing* 9, 1 (2016), 51–65.
[33] Sheldon Cohen, Ronald C Kessler, and Lynn Underwood Gordon. 1997. *Measuring stress: A guide for health and social scientists*. Oxford University Press on Demand.
[34] Cary L Cooper and Sue Cartwright. 1997. An intervention strategy for workplace stress. *Journal of psychosomatic research* 43, 1 (1997), 7–16.
[35] Neil E Cotter. 1990. The Stone-Weierstrass theorem and its application to neural networks. *IEEE transactions on neural networks* 1, 4 (1990), 290–295.
[36] Alberto de Santos Sierra, Carmen Sánchez Ávila, Gonzalo Bailador del Pozo, and Javier Guerra Casanova. 2011. Stress detection by means of stress physiological template. In *2011 Third World Congress on Nature and Biologically Inspired Computing*. IEEE, 131–136.
[37] Debidatta Dwibedi, Yusuf Aytar, Jonathan Tompson, Pierre Sermanet, and Andrew Zisserman. 2020. Counting Out Time: Class Agnostic Video Repetition Counting in the Wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10387–10396.
[38] Jonathan Foote. 1999. Visualizing music and audio using self-similarity. In *Proceedings of the seventh ACM international conference on Multimedia (Part 1)*. 77–80.
[39] Jonathan Foote and Matthew Cooper. 2001. Visualizing Musical Structure and Rhythm via Self-Similarity.. In *ICMC*, Vol. 1. Citeseer, 423–430.
[40] Jonathan T Foote and Matthew L Cooper. 2003. Media segmentation using self-similarity decomposition. In *Storage and Retrieval for Media Databases 2003*, Vol. 5021. International Society for Optics and Photonics, 167–175.
[41] Daniel Fuller, Emily Colwell, Jonathan Low, Kassia Orychock, Melissa Ann Tobin, Bo Simango, Richard Buote, Desiree Van Heerden, Hui Luan, Kimberley Cullen, et al. 2020. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: Systematic review. *JMIR mHealth and uHealth* 8, 9 (2020), e18694.
[42] Enrique Garcia-Ceja, Venet Osmani, and Oscar Mayora. 2015. Automatic stress detection in working environments from smartphones' accelerometer data: a first step. *IEEE journal of biomedical and health informatics* 20, 4 (2015), 1053–1060.
[43] Miguel A García-González, Ariadna Argelagós-Palau, Mireya Fernández-Chimeno, and Juan Ramos-Castro. 2013. A comparison of heartbeat detectors for the seismocardiogram. In *Computing in Cardiology 2013*. IEEE, 461–464.
[44] Renee Garett, Sam Liu, and Sean D Young. 2017. A longitudinal analysis of stress among incoming college freshmen. *Journal of American college health* 65, 5 (2017), 331–338.
[45] Z Ge, PWC Prasad, N Costadopoulos, Abeer Alsadoon, AK Singh, and A Elchouemi. 2016. Evaluating the accuracy of wearable heart rate monitors. In *2016 2nd International Conference on Advances in Computing, Communication, & Automation (ICACCA)(Fall)*. IEEE, 1–6.
[46] Konstantinos Georgiou, Andreas V Larentzakis, Nehal N Khamis, Ghadah I Alsuhaibani, Yasser A Alaska, and Elias J Giallafos. 2018. Can wearable devices accurately measure heart rate variability? A systematic review. *Folia medica* 60, 1 (2018), 7–20.
[47] Dimitris Giakoumis, Anastasios Drosou, Pietro Cipresso, Dimitrios Tzovaras, George Hassapis, Andrea Gaggioli, and Giuseppe Riva. 2012. Using activity-related behavioural features towards more effective automatic stress detection. *PloS one* 7, 9 (2012), e43571.
[48] G Giannakakis, Matthew Pediaditis, Dimitris Manousos, Eleni Kazantzaki, Franco Chiarugi, Panagiotis G Simos, Kostas Marias, and Manolis Tsiknakis. 2017. Stress and anxiety detection using facial cues from videos. *Biomedical Signal Processing and Control* 31 (2017), 89–101.
[49] Francesca Gino. 2016. Are you too stressed to be productive? Or not stressed enough. *Harvard Business Review* (2016).
[50] Martin Gjoreski, Hristijan Gjoreski, Mitja Luštrek, and Matjaž Gams. 2016. Continuous stress detection using a wrist device: in laboratory and real life. In *proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing: Adjunct*. 1185–1193.
[51] Martin Gjoreski, Mitja Luštrek, Matjaž Gams, and Hristijan Gjoreski. 2017. Monitoring stress with a wrist device using context. *Journal of biomedical informatics* 73 (2017), 159–170.
[52] Unsoo Ha, Salah Assana, and Fadel Adib. 2020. Contactless seismocardiography via deep learning radars. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*. 1–14.
[53] George Halkos and Dimitrios Bousinakis. 2010. The effect of stress and satisfaction on productivity. *International Journal of Productivity and Performance Management* (2010).

[54] Martica Hall, Raymond Vasko, Daniel Buysse, Hernando Ombao, Qingxia Chen, J David Cashmere, David Kupfer, and Julian F Thayer. 2004. Acute stress affects heart rate variability during sleep. *Psychosomatic medicine* 66, 1 (2004), 56–62.

[55] Anne Harrington. 1999. *The placebo effect: An interdisciplinary exploration*. Vol. 8. Harvard University Press.

[56] Sandra G Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage publications Sage CA: Los Angeles, CA, 904–908.

[57] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.

[58] Marit Hauschildt, Maarten JV Peters, Steffen Moritz, and Lena Jelinek. 2011. Heart rate variability in response to affective scenes in posttraumatic stress disorder. *Biological Psychology* 88, 2-3 (2011), 215–222.

[59] Jennifer A Healey and Rosalind W Picard. 2005. Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on intelligent transportation systems* 6, 2 (2005), 156–166.

[60] Brian Heater. 2021. Google's Soli radar returns to track sleep on the new Nest Hub. Tech Crunch. https://techcrunch.com/2021/03/16/googles-soli-returns-to-track-sleep-on-the-new-nest-hub/.

[61] Keith C Hendy, Kevin M Hamilton, and Lois N Landry. 1993. Measuring subjective workload: when is one scale better than many? *Human Factors* 35, 4 (1993), 579–601.

[62] Tin Kam Ho. 1995. Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition*, Vol. 1. IEEE, 278–282.

[63] Karen Hovsepian, Mustafa Al'Absi, Emre Ertin, Thomas Kamarck, Motohiro Nakajima, and Santosh Kumar. 2015. cStress: towards a gold standard for continuous stress assessment in the mobile environment. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*. 493–504.

[64] Osama Talaat Ibrahim, Walid Gomaa, and Moustafa Youssef. 2019. Crosscount: A deep learning system for device-free human counting using wifi. *IEEE Sensors Journal* 19, 21 (2019), 9921–9928.

[65] CV Jensen and T Bendix. 1992. Spontaneous movements with various seated-workplace adjustments. *Clinical Biomechanics* 7, 2 (1992), 87–90.

[66] Giorgos Karvounas, Iason Oikonomidis, and Antonis Argyros. 2019. ReActNet: Temporal Localization of Repetitive Activities in Real-World Videos. *arXiv preprint arXiv:1910.06096* (2019).

[67] Shwetambara Kekade, Chung-Ho Hseieh, Md Mohaimenul Islam, Suleman Atique, Abdulwahed Mohammed Khalfan, Yu-Chuan Li, and Shabbir Syed Abdul. 2018. The usefulness and actual use of wearable devices among the elderly population. *Computer methods and programs in biomedicine* 153 (2018), 137–159.

[68] M Kellmann. 2010. Preventing overtraining in athletes in high-intensity sports and stress/recovery monitoring. *Scandinavian journal of medicine & science in sports* 20 (2010), 95–102.

[69] Hye-Geum Kim, Eun-Jin Cheon, Dai-Seg Bai, Young Hwan Lee, and Bon-Hoon Koo. 2018. Stress and heart rate variability: a meta-analysis and review of the literature. *Psychiatry investigation* 15, 3 (2018), 235.

[70] Jonghwa Kim and Elisabeth André. 2008. Emotion recognition based on physiological changes in music listening. *IEEE transactions on pattern analysis and machine intelligence* 30, 12 (2008), 2067–2083.

[71] Zachary D King, Judith Moskowitz, Begum Egilmez, Shibo Zhang, Lida Zhang, Michael Bass, John Rogers, Roozbeh Ghaffari, Laurie Wakschlag, and Nabil Alshurafa. 2019. Micro-stress EMA: A passive sensing framework for detecting in-the-wild stress in pregnant mothers. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 3, 3 (2019), 1–22.

[72] Tina HP Kowalski and Wendy Loretto. 2017. Well-being and HRM in the changing workplace.

[73] Hindra Kurniawan, Alexandr V Maslov, and Mykola Pechenizkiy. 2013. Stress detection from speech and galvanic skin response signals. In *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*. IEEE, 209–214.

[74] Susan Levenstein, Cosimo Prantera, Vilma Varvo, Maria L Scribano, Eva Berto, Carlo Luzi, and Arnaldo Andreoli. 1993. Development of the Perceived Stress Questionnaire: a new tool for psychosomatic research. *Journal of psychosomatic research* 37, 1 (1993), 19–32.

[75] Alexandros Liapis, Christos Katsanos, Dimitris G Sotiropoulos, Nikos Karousos, and Michalis Xenos. 2017. Stress in interactive applications: analysis of the valence-arousal space based on physiological signals and self-reported data. *Multimedia Tools and Applications* 76, 4 (2017), 5051–5071.

[76] Andy Liaw, Matthew Wiener, et al. 2002. Classification and regression by randomForest. *R news* 2, 3 (2002), 18–22.

[77] Yu-Ru Lin, Hari Sundaram, Yun Chi, Junichi Tatemura, and Belle L Tseng. 2007. Splog detection using self-similarity analysis on blog temporal dynamics. In *Proceedings of the 3rd international workshop on Adversarial information retrieval on the web*. 1–8.

[78] Ivan Liu, Shiguang Ni, and Kaiping Peng. 2020. Happiness at Your Fingertips: Assessing Mental Health with Smartphone Photoplethysmogram-Based Heart Rate Variability Analysis. *Telemedicine and e-Health* 26, 12 (2020), 1483–1491.

[79] Sonia J Lupien and F Seguin. 2013. How to measure stress in humans. *Centre for Studies in Human Stress* (2013).

[80] Choubeila Maaoui, Alain Pruski, and Faiza Abdat. 2008. Emotion Recognition for hHman-Machine Communication. In *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 1210–1215.

[81] Alberto Machado, Antonio J Herrera, Rocío M de Pablos, Ana María Espinosa-Oliva, Manuel Sarmiento, Antonio Ayala, José Luis Venero, Martiniano Santiago, Ruth F Villarán, María José Delgado-Cortés, et al. 2014. Chronic stress as a risk factor for Alzheimer's disease. *Reviews in the Neurosciences* 25, 6 (2014), 785–804.
[82] M Loredana Marcovecchio and Francesco Chiarelli. 2012. The effects of acute and chronic stress on diabetes control. *Science signaling* 5, 247 (2012), pt10–pt10.
[83] Mathworks. [n.d.]. Findpeaks Function. https://www.mathworks.com/help/signal/ref/findpeaks.html
[84] Daniel McDuff, Amy Karlson, Ashish Kapoor, Asta Roseway, and Mary Czerwinski. 2012. AffectAura: an intelligent system for emotional memory. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 849–858.
[85] Paolo Melillo, Marcello Bracale, and Leandro Pecchia. 2011. Nonlinear Heart Rate Variability features for real-life stress detection. Case study: students under stress due to university examination. *Biomedical engineering online* 10, 1 (2011), 1–13.
[86] Varun Mishra, Gunnar Pope, Sarah Lord, Stephanie Lewia, Byron Lowens, Kelly Caine, Sougata Sen, Ryan Halter, and David Kotz. 2018. The case for a commodity hardware solution for stress detection. In *Proceedings of the 2018 ACM international joint conference and 2018 international symposium on pervasive and ubiquitous computing and wearable computers*. 1717–1728.
[87] Varun Mishra, Sougata Sen, Grace Chen, Tian Hao, Jeffrey Rogers, Ching-Hua Chen, and David Kotz. 2020. Evaluating the Reproducibility of Physiological Stress Detection Models. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (2020), 1–29.
[88] Ranjita Misra and Linda G Castillo. 2004. Academic stress among college students: Comparison of American and international students. *International Journal of stress management* 11, 2 (2004), 132.
[89] Miles Y Muraoka, John G Carlson, and Claude M Chemtob. 1998. Twenty-four-hour ambulatory blood pressure and heart rate monitoring in combat-related posttraumatic stress disorder. *Journal of Traumatic Stress: Official Publication of The International Society for Traumatic Stress Studies* 11, 3 (1998), 473–484.
[90] Armando Muscariello, Guillaume Gravier, and Frédéric Bimbot. 2011. Towards robust word discovery by self-similarity matrix comparison. In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 5640–5643.
[91] Helga Þórarinsdóttir, Lars Vedel Kessing, and Maria Faurholt-Jepsen. 2017. Smartphone-based self-assessment of stress in healthy adult individuals: a systematic review. *Journal of medical Internet research* 19, 2 (2017), e41.
[92] Aoife O'Donovan, A. Janet Tomiyama, Jue Lin, Eli Puterman, Nancy E. Adler, Margaret Kemeny, Owen M. Wolkowitz, Elizabeth H. Blackburn, and Elissa S. Epel. 2012. Stress appraisals and cellular aging: A key role for anticipatory threat in the relationship between psychological stress and telomere length. *Brain, Behavior, and Immunity* 26, 4 (2012), 573–579. https://doi.org/10.1016/j.bbi.2012.01.007
[93] K Palanisamy, M Murugappan, and S Yaacob. 2013. Multiple physiological signal-based human stress identification using non-linear classifiers. *Elektronika ir elektrotechnika* 19, 7 (2013), 80–85.
[94] HyeYoun Park, Sohee Oh, Yumi Noh, Ju Young Kim, and Jeong-Hyun Kim. 2018. Heart rate variability as a marker of distress and recovery: the effect of brief supportive expressive group therapy with mindfulness in cancer patients. *Integrative cancer therapies* 17, 3 (2018), 825–831.
[95] Linda Gay Peterson and Lori Pbert. 1992. Effectiveness of a meditation-based stress reduction program in the treatment of anxiety disorders. *Am J Psychiatry* 149, 7 (1992), 936–943.
[96] Rosalind W. Picard, Elias Vyzas, and Jennifer Healey. 2001. Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE transactions on pattern analysis and machine intelligence* 23, 10 (2001), 1175–1191.
[97] Monika Pobiruchin, Julian Suleder, Richard Zowalla, and Martin Wiesner. 2017. Accuracy and adoption of wearable technology used by active citizens: a marathon event field study. *JMIR mHealth and uHealth* 5, 2 (2017), e24.
[98] Octavian Postolache, Pedro Silva Girão, Eduardo Pinheiro, and Gabriela Postolache. 2010. Unobtrusive and non-invasive sensing solutions for on-line physiological parameters monitoring. In *Wearable and Autonomous Biomedical Devices and Systems for Smart Environment*. Springer, 277–314.
[99] Samma Faiz Rasool, Mansi Wang, Yanping Zhang, and Madeeha Samma. 2020. Sustainable work performance: the roles of workplace violence and occupational stress. *International journal of environmental research and public health* 17, 3 (2020), 912.
[100] Susana Rubio, Eva Díaz, Jesús Martín, and José M Puente. 2004. Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. *Applied psychology* 53, 1 (2004), 61–86.
[101] Takuya Sakamoto, Ryohei Imasaka, Hirofumi Taki, Toru Sato, Mototaka Yoshioka, Kenichi Inoue, Takeshi Fukuda, and Hiroyuki Sakai. 2015. Feature-based correlation and topological similarity for interbeat interval estimation using ultrawideband radar. *IEEE Transactions on Biomedical Engineering* 63, 4 (2015), 747–757.
[102] Hillol Sarker, Matthew Tyburski, Md Mahbubur Rahman, Karen Hovsepian, Moushumi Sharmin, David H Epstein, Kenzie L Preston, C Debra Furr-Holden, Adam Milam, Inbal Nahum-Shani, et al. 2016. Finding significant stress episodes in a discontinuous time series of rapidly varying mobile sensor data. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 4489–4501.
[103] Urte Scholz, Roberto La Marca, Urs M Nater, Ingo Aberle, Ulrike Ehlert, Rainer Hornung, Mike Martin, and Matthias Kliegel. 2009. Go no-go performance under psychosocial stress: Beneficial effects of implementation intentions. *Neurobiology of learning and memory* 91, 1 (2009), 89–92.

[104] Daniela Schoofs, Diana Preuß, and Oliver T Wolf. 2008. Psychosocial stress induces working memory impairments in an n-back paradigm. *Psychoneuroendocrinology* 33, 5 (2008), 643–653.
[105] Ssang-Hee Seo, Jung-Tae Lee, and Marius Crisan. 2010. Stress and EEG. *Convergence and hybrid information technologies* 27 (2010).
[106] Mohsin Shah, Shahid Hasan, Samina Malik, and Chandrashekhar T Sreeramareddy. 2010. Perceived stress, sources and severity of stress among medical undergraduates in a Pakistani medical school. *BMC medical education* 10, 1 (2010), 1–8.
[107] Yuan Shi, Minh Hoai Nguyen, Patrick Blitz, Brian French, Scott Fisk, Fernando De la Torre, Asim Smailagic, Daniel P Siewiorek, Mustafa al'Absi, Emre Ertin, et al. 2010. Personalized stress detection from physiological measurements. In *International symposium on quality of life technology*. 28–29.
[108] Mihoko Shimano, Takahiro Okabe, Imari Sato, and Yoichi Sato. 2010. Video temporal super-resolution based on self-similarity. In *Asian Conference on Computer Vision*. Springer, 93–106.
[109] Nataša Šimić and Ilija Manenica. 2012. Exam experience and some reactions to exam stress. *Human physiology* 38, 1 (2012), 67–72.
[110] Gottfried Spangler. 1997. Psychological and physiological responses during an exam and their relation to personality characteristics. *Psychoneuroendocrinology* 22, 6 (1997), 423–441.
[111] Gijsbert Stoet. 2010. PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior research methods* 42, 4 (2010), 1096–1104.
[112] Gijsbert Stoet. 2017. PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology* 44, 1 (2017), 24–31.
[113] Joachim Taelman, Tine Adriaensen, Caroline van Der Horst, Torsten Linz, and Arthur Spaepen. 2007. Textile integrated contactless EMG sensing for stress analysis. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 3966–3969.
[114] Gustavo E. Tafet and Renato Bernardini. 2003. Psychoneuroendocrinological links between chronic stress and depression. *Progress in Neuro-Psychopharmacology and Biological Psychiatry* 27, 6 (2003), 893–903. https://doi.org/10.1016/S0278-5846(03)00162-3
[115] Chek Tien Tan, Tuck Wah Leong, and Songjia Shen. 2014. Combining think-aloud and physiological data to understand video game experiences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 381–390.
[116] Karl Halvor Teigen. 1994. Yerkes-Dodson: A law for all seasons. *Theory & Psychology* 4, 4 (1994), 525–547.
[117] David Tse and Pramod Viswanath. 2005. *Fundamentals of wireless communication*. Cambridge university press.
[118] V Vanitha and P Krishnan. 2016. Real time stress detection system based on EEG signals. (2016).
[119] Alexandros N Vgontzas, Constantine Tsigos, Edward O Bixler, Constantine A Stratakis, Keith Zachman, Anthony Kales, Antonio Vela-Bueno, and George P Chrousos. 1998. Chronic insomnia and activity of the stress system: a preliminary study. *Journal of psychosomatic research* 45, 1 (1998), 21–31.
[120] Elena Vildjiounaite, Johanna Kallio, Vesa Kyllönen, Mikko Nieminen, Ilmari Määttänen, Mikko Lindholm, Jani Mäntyjärvi, and Georgy Gimel'farb. 2018. Unobtrusive stress detection on the basis of smartphone usage data. *Personal and Ubiquitous Computing* 22, 4 (2018), 671–688.
[121] Wei Wang, Alex X Liu, and Muhammad Shahzad. 2016. Gait recognition using wifi signals. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 363–373.
[122] Chang Zhi Wei. 2013. Stress emotion recognition based on RSP and EMG signals. In *Advanced Materials Research*, Vol. 709. Trans Tech Publ, 827–831.
[123] I-Chieh Wei, Chih-Wei Wu, and Li Su. 2019. Generating Structured Drum Pattern Using Variational Autoencoder and Self-similarity Matrix.. In *ISMIR*. 847–854.
[124] Jacqueline Wijsman, Bernard Grundlehner, Hao Liu, Julien Penders, and Hermie Hermens. 2013. Wearable physiological sensors reflect mental stress state in office-like situations. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*. IEEE, 600–605.
[125] Min Wu, Hong Cao, Hai-Long Nguyen, Karl Surmacz, and Caroline Hargrove. 2015. Modeling perceived stress via HRV and accelerometer sensor streams. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 1625–1628.
[126] Chang-Ming Yang, Chih-Chung Wu, Chun-Mei Chou, and Ching-Wen Yang. 2010. Textile-based breath-sensing belt. In *2010 Digest of Technical Papers International Conference on Consumer Electronics (ICCE)*. IEEE, 11–12.
[127] Youwei Zeng, Dan Wu, Jie Xiong, Jinyi Liu, Zhaopeng Liu, and Daqing Zhang. 2020. MultiSense: Enabling multi-person respiration sensing with commodity wifi. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 3 (2020), 1–29.
[128] Mingmin Zhao, Fadel Adib, and Dina Katabi. 2016. Emotion recognition using wireless signals. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*. 95–108.
[129] Mingmin Zhao, Tianhong Li, Mohammad Abu Alsheikh, Yonglong Tian, Hang Zhao, Antonio Torralba, and Dina Katabi. 2018. Through-wall human pose estimation using radio signals. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 7356–7365.

[130] Mingmin Zhao, Yingcheng Liu, Aniruddh Raghu, Tianhong Li, Hang Zhao, Antonio Torralba, and Dina Katabi. 2019. Through-wall human mesh recovery using radio signals. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 10113–10122.

[131] Bin Zheng, Xianta Jiang, Geoffrey Tien, Adam Meneghetti, O Neely M Panton, and M Stella Atkins. 2012. Workload assessment of surgeons: correlation between NASA TLX and blinks. *Surgical endoscopy* 26, 10 (2012), 2746–2750.