Wi-Watch: WiFi-based Vigilant-Activity Recognition for Ship Bridge Watchkeeping Officers

Danei Gong⋆, Kezhong Liu⋆, *Member, IEEE,* Dashuai Pei,
Haoran Zhang, Shengkai Zhang, *Member, IEEE,* Mozi Chen*, *Member, IEEE*

Abstract—To ensure the safety of marine traffic, ship watchkeeping officers must maintain a high level of vigilance during their watchkeeping period. Although advanced driving assistance systems are available that can effectively measure driver alertness and provide early warnings before any hazardous maneuvers, such systems are primarily designed for road vehicles and are considerably different from those used when operating a marine vehicle. In this paper, we propose to use fine-grained channel state information (CSI) obtained using commercial off-the-shelf wireless fidelity (WiFi) to track an officer-on-watch's vigilant activities and determine, in real time, whether the ship officer is adhering to safety guidelines. We developed a CSI path model and a 2D multiple signal classification (MUSIC) algorithm to estimate signal propagation path parameters, such as angle of arrival and Doppler shift, in a low signal-to-noise ratio (SNR) ship environment. We then developed CSI velocity and activity models based on finer-grained signal parameters, using wavelet transforms and deep learning techniques to determine bridge-officer activities. Finally, we developed a watchkeeping vigilance evaluation module to determine whether a bridge officer was complying with safe driving guidelines during their watchkeeping. Our proposed system was implemented on a commercial WiFi platform, and we extensively evaluated the system on an actual passenger ship. Our proposed system achieved accuracies of 95.4% and 93.8% for tracking walking movements and recognizing careless activities, respectively.

Index Terms—Ship watchkeeping, vigilance detection, human activities recognition, wireless sensing

I. INTRODUCTION

WATCHKEEPING is one of the most critical responsibilities of ship bridge officers-on-watch (OOWs) during maritime navigation, and it plays a vital role in improving ship traffic safety. Low vigilance during watchkeeping may lead to marine vehicle grounding or collision, resulting in significant loss of human life and/or property damage [1]. According to the annual report on marine casualties and incidents issued by the European Maritime Safety Agency, 71% of the common contributory factors for the 1,170 cited

K. Liu is with the School of Navigation, Wuhan University of Technology, Wuhan 430063, China; the National Engineering Research Center for Water Transport Safety, Wuhan 430063, China; and the Hubei Key Laboratory of Inland Shipping Technology, Wuhan 430063, China (e-mail: kzliu@whut.edu.cn).

D. Gong, D. Pei, M. Chen, and H. Zhang are with the School of Navigation, Wuhan University of Technology, Wuhan 430063, China (e-mail: dank, Pei.Dashuai, chenmz, haoran.zhang@whut.edu.cn).

S. Zhang is with the School of Information Engineering, Wuhan University of Technology, Wuhan 430063, China (e-mail: shengkai@whut.edu.cn).

⋆ Danei Gong and Kezhong Liu are co-first authors.

* Corresponding author: Mozi Chen (chenmz@whut.edu.cn).

The datasets and code used for this research are available. URL: https://gitee.com/ybw272736225/vimmfi_datasets.
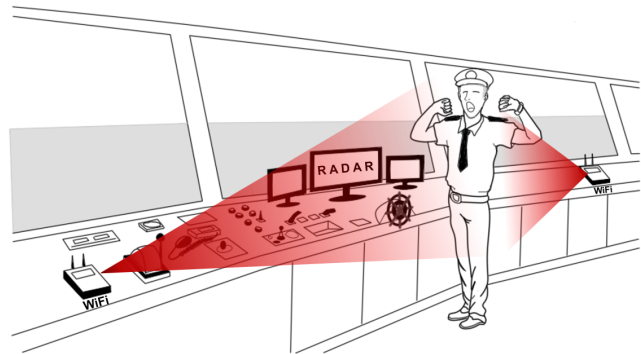
Fig. 1: A illustration of OOW activity monitoring using WiFi signal. The activities can be captured in a device-free manner based on existing WiFi infrastructures.

collisions or grounding accidents were mainly due to poor lookout and poor radar use [2]. In a recent series of notable accidents, the common features included fatigued officers, one-man bridge operation at night, and no watch alarms. Most of such errors arise from a driver's failure to comply with vigilant watchkeeping guidelines [3], including manual distractions (e.g., when the driver leaves the bridge, falls asleep, or becomes fatigued), visual distractions (e.g., when the driver uses a smartphone or a reading book), less vigilance to the sailing environment (e.g., when the driver fails to notice an oncoming ship by radar or binoculars), and incorrect steering techniques [4–6]. Many aquatic traffic accidents can be prevented if such careless driving behaviors are automatically detected in real time.

The use of artificial intelligence technology to detect dangerous driving behaviors in real time and evaluate driving safety risks has recently attracted increasing attention. Unfortunately, because such studies have mainly focused on road vehicles such as cars and trucks, the study findings cannot be directly applied to monitoring marine vehicle drivers. Existing driver-monitoring systems can be categorized into device-based or device-free systems. Device-based systems mainly use wearable devices to monitor driving activities. For example, previous studies have used inertial measurement units (IMU), commonly found on smartwatches, to detect dangerous steering wheel activities [7]. Another method for detecting a driver's sustained-attention behavior is to analyze brain dynamics using electroencephalography (EEG) data [8]. However, such solutions only work while the driver is wearing a smartwatch on their wrist or an EEG signal collector on

their head, which limits the extensive real-world application of such systems to ships. Device-free methods are mostly based on computer vision [9], acoustic signals [10], and infrared light [11]. Among them, computer-vision-based systems use cameras installed around the driver's seat to capture videos of the driver's face and detect visual distractions [12]. However, the movement space required for a ship officer to perform driving activities is much larger than that required for driving vehicles. Therefore, it is challenging to maintain long high-resolution videos of ship officers' faces. Although depth cameras, such as Microsoft Kinect [13], and infrared LEDs, can be used to detect the presence and movement of ship officers, ship watchkeeping officers have to perform a wide range of activities to complete complex tasks during their long shifts. Additionally, surveillance videos frequently cause concerns regarding privacy as cameras capture individuals' faces, attire, and private behaviors [14] [15]. Prolonged exposure to these videos can adversely affect the psychological and emotional state of employees, with potential negative consequences on their work [16] [17]. Therefore, there is a need to develop reliable, robust, and low-cost technologies to monitor the various activities of ship officers.

With the rapid development of mobile computing and wireless infrastructures, using WiFi signals to detect human activities has shown great potential. This sensing approach can be achieved by reusing existing WiFi devices, which have the characteristics of ubiquity and low cost. The intuition behind this concept is that moving-object-induced interference, including WiFi signal absorption, reflection, and refraction, can largely change the multipath propagation between transceivers in an indoor environment. Moreover, it has been shown that environmental changes, such as human presence and movement, can affect the propagation of wireless signals and change wireless channels. Specifically, physical-layer channel state information (CSI), which can be obtained from commercial WiFi devices, can reveal multipath wireless channels at the granularity of orthogonal frequency-division multiplexing (OFDM) subcarriers [18]. By exploiting CSI amplitude and phase information and learning the characteristics of their variations, device-free human activity recognition can be achieved, including localization and tracking [19][20], fall detection [21], gesture recognition [22], presence detection [23], pose estimation [24], state estimation [25], and vital-sign monitoring [26].

This paper presents Wi-Watch, a WiFi-based shipbridge watchkeeping vigilance tracking system. We propose the Wi-Watch system, which could function not just as an auxiliary to the video-based OOW vigilance tracking system, but also markedly diminish the privacy implications led by camera usage. An illustration of Wi-Watch is shown in Fig. 1. Intuitively, most inattentive OOW behaviors such as sleeping, reading, and using smartphones are usually static on the bridge, which negligibly interferes with indoor WiFi channels, while lookout behaviors such as using radar and checking the sea route will significantly change the CSI owing to OOW movement. To accurately track OOW movements using CSI, we first investigated various wireless-signal-processing algorithms and found that the MUSIC algorithm [27], which can jointly estimate the angle of arrival (AoA) and time-of-flight (ToF) based on CSI diversity across multi-antennas in multiple-input multiple-output (MIMO) systems, is promising for tracking bridge officers. However, low SNR environments and vibrating background reflections can considerably degrade the performance of the MUSIC algorithm in mobile ships [28] [29]. To overcome these challenges, we first adopted a CSI path model and a 2D MUSIC algorithm to estimate the signal-propagation path parameters (AoA and Doppler shift) in a low-SNR environment instead of ToF. Then, we resolved the multipath at a much finer-grained resolution to extract the OOW's reflected signals from background vibrations through a proposed foreground extraction algorithm. The extracted fine-grained OOW reflected path parameters (i.e., Doppler shift and AoA) were then used to evaluate the OOW's vigilance according to the OOW's on-duty rules. The vigilance-tracking-system design includes four modules: *Activity state determination*, *walking velocity estimation*, *in-place activity recognition*, and *vigilance evaluation*. We implemented our system on a commercial WiFi platform and extensively evaluated the system on an actual ship. We demonstrated that our system achieved accuracies of 95.4% and 93.8% for recognizing activity state and recognizing careless activities, respectively. The contributions of this study are as follows:

- We developed a WiFi-based method of recognizing the vigilant behaviors of ship OOWs, which can preserve officer privacy and alert officers when they are less vigilant during watchkeeping.
- A CSI-path model was adopted and a MUSIC-based path parameter estimation method was proposed to obtain a Doppler-AoA spectrum of the target reflection path to describe the OOW's activity.
- We designed an attention-based long short-term memory (LSTM) deep learning architecture to recognize the key OOW lookout activities by converting the CSI segments into 2D-image-based input representations.
- We implemented the Wi-Watch system in a mobile ship environment with commercial WiFi devices, and the experimental results showed that the system achieved an overall recognition accuracy of 93.8%.

This paper serves as an expanded version of our previous publication [30] presented at The 2023 International Conference on Marine Equipment & Technology and Sustainable Development (Publisher: Springer Nature, Pages: 970-977, Year: 2023).

II. MODELING VIGILANCE

We first introduce the duties OOWs perform during their watch and illustrate several vigilant and non vigilant activities decomposed into detailed actions. Then, we propose our basic idea of tracking OOW vigilance with WiFi signals and show our motivational experiments.

A. OOW Duties

OOWs are the deck officers responsible for watching and navigating on the bridge. They continuously monitor the sea
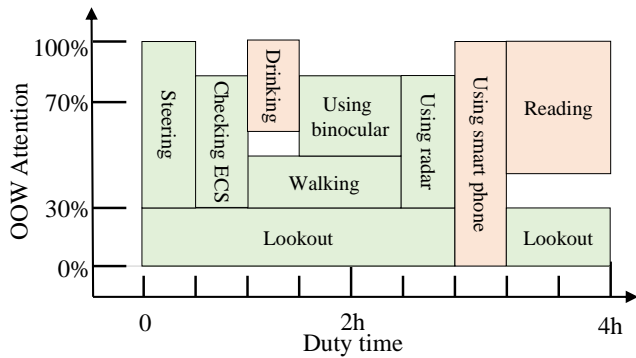
Fig. 2: An illustration of the OOW duty cycle with the duration and attention occupation of each action. All actions are marked with two colors to indicate vigilant and non-vigilant actions.
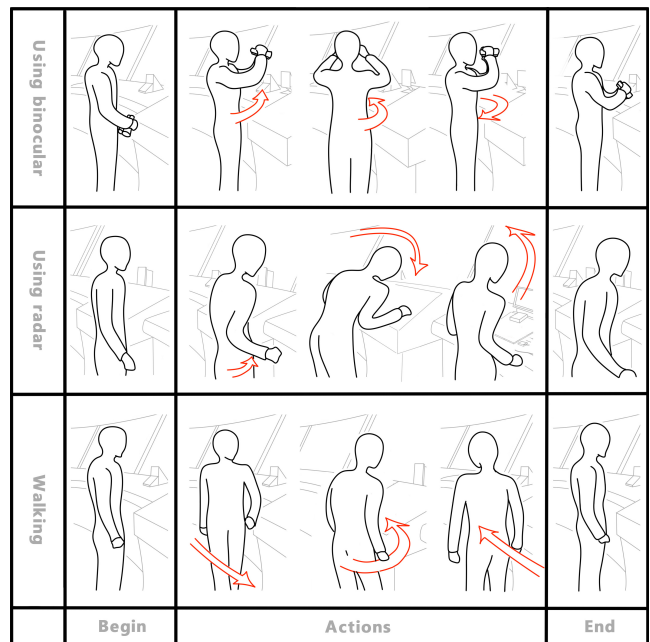

Fig. 3: OOW activities sketches. Here, we show three representative activities of OOW during his watch: using radar, using binoculars, and walking, and we decompose them into three phases.

surface and report any hazards that may obstruct navigation and damage the ship. According to the International Regulations for Preventing Collisions at Sea (COLREGS), surveillance requires uninterrupted attention to the ship's navigation to inform about other ships, shipwrecks, wrecks, floating objects, etc. The International Convention on Standards of Training, Certification, and Watchkeeping for Seafarers (STCW) stipulates that: Article 5 of the "1972 International Regulations for Preventing Collisions at Sea" shall be observed at all times, and the following objectives shall be achieved: 1) Use sight, hearing, and all other available methods to maintain a continuous alarm state; 2) Comprehensively judge situations and dangers that endanger the safety of navigation, such as collision and grounding; 3) Detect ships, airplanes, ship victims, shipwrecks, and other navigational obstacles.

Fig. 2 shows the OOW duty cycle and indicates the duration and attention occupation of each watchkeeping action. The OOW duties are as follows: 1) Lookout: The officer should concentrate on maintaining a regular lookout continuously and shall not engage in or assign other tasks that will affect the lookout; 2) Using radar: Using radar to watch is crucial to the safety and must be maintained on the stipulated frequencies, as per the regulations. 3) Compare the electronic charting system (ECS): The OOW must check the position plotted by the outgoing OOW and not depend entirely on the information displayed on the chart. 4) The responsibilities of watchmen and helmsmen should be separated, and the steering helmsmen should not be regarded as watchmen (except on small boats). 5) Walking rounds of the bridge: Soon after handing over the watch, the relieved OOW may take a round of the ship to ascertain that fire safety is maintained, there are no signs of a breach, nothing unusual, and no unsecured articles in the accommodation. 6) Reading log: The OOW must read any log entries made by the outgoing OOW before he leaves the bridge.

B. OOW Vigilant-activities

It is well known that long shifts and distraction from others lead to watchkeeper non-vigilance, and the inability to strictly follow these rules leads to frequent accidents. Here, we define some of the most common safe and unsafe OOW actions. As shown in Fig. 2, several positive and negative actions are associated with bridge watchkeeping. Positive actions include *using radar*, *using binoculars*, and *using the steering rudder* and can indicate that the OOW's attention is completely focused on ship safety. In contrast, some negative actions including *using smartphones*, *drinking water*, and *reading/writing* may distract the OOW's attention away from conducting a proper lookout. An illustration of each OOW's actions is presented in Fig. 3. We introduce the typical steps involved in using radar as an example. Usually, radar usage can be divided into the beginning, operating, and ending phases. After walking over to and standing before the radar, the OOW usually bends over and raises their hand to operate the radar. While checking the radar, the OOW's head and radar-operating hand would swing slightly and intermittently. After that, the OOW stands up and leaves to go perform other operations. These actions can be further decomposed into six detailed steps: 1) walking, 2) standing, 3) stopping, 4) raising hands, 5) swinging hand and head, and 6) standing up.

C. Vigilance-Tracking Intuition

Our work aims to evaluate the vigilance of the OOW by extracting the WiFi channel parameters and continuously recognizing the OOW's real-time actions. Instead of video-based watchkeeper-monitoring systems, the WiFi-based activity-recognition system can preserve crewmembers' privacy without using a camera and can be deployed using the existing onboard WiFi infrastructure without incurring additional costs. The basic idea behind tracking the OOW's vigilance is to recognize the specific positive lookout actions by measuring the WiFi channel parameters. It is well known
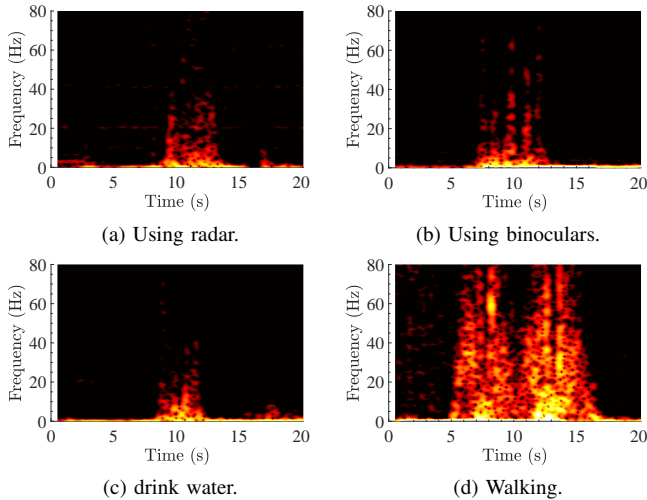
Fig. 4: Four different human activities have different effects on WiFi CSI.

that in indoor environments, wireless signals are usually propagated from the source to the receiver by reflecting off objects such as walls and furniture, which is called the multipath effect. When the OOW performs a gesture, their body parts (e.g., arms, legs, and torso) move at different positions. The wireless signals reflected off the body parts would propagate over various distances and be superposed at the receiver to form the corresponding CSI profile. By extracting the WiFi propagation distance (i.e., ToF), AoA, and Doppler velocity of the reflected WiFi-signal paths using a pair of APs and an array of receivers, we can identify the OOW's actions.

To facilitate the OOW's activity recognition and vigilance identification using WiFi signals, we exploit CSI, the frequency description of the wireless channel, and extract the wireless path parameters using signal processing techniques such as the MUSIC algorithm. With the IEEE 802.11n MIMO system, fine-grained CSI measurements from multiple antennas can be extracted by leveraging commercial WiFi network interface cards (NICs) such as an Intel® 5300 and an Atheros 9390 with public open-source tools [31] [32]. The CSI tools utilize a channel-sounding mechanism in WiFi 802.11n/g to report the CSI from the receiver for every received packet. For example, the Intel 5300 NIC reports CSI for 30 groups of subcarriers, spread evenly among the 56 subcarriers of a 20 MHz channel or the 114 carriers in a 40 MHz channel. The channel matrix (X) obtained from 3 antennas can be presented as follows:

$$\text{CSI matrix} = \begin{bmatrix} \text{csi}_{1,1} & \text{csi}_{1,2} & \cdots & \text{csi}_{1,30} \\ \text{csi}_{2,1} & \text{csi}_{2,2} & \cdots & \text{csi}_{2,30} \\ \text{csi}_{3,1} & \text{csi}_{3,2} & \cdots & \text{csi}_{3,30} \end{bmatrix}. \quad (1)$$

Each entry is a complex number, where $\text{csi}_{m,n}$ denotes the CSI value for the m-th antenna and the n-th subcarrier. $|\text{csi}_{m,n}|$ is the amplitude, and $\angle \text{csi}_{m,n}$ denotes the phase.

Fig. 4 presents the results of a motivational experiment illustrating the distinguished effects of four human activities on WiFi CSI: using a radar, using binoculars, drinking water, and walking. We aimed to demonstrate the ability to recognize

human activities through the distinct profiles of CSI signal amplitudes. Initially, we randomly selected a subcarrier with human activity from the CSI with 30 subcarriers and calculated its amplitude. Subsequently, we performed the short-time Fourier transform (STFT) on the amplitude signal to obtain its spectrogram. This step helped us visualize the energy of each frequency component over time, and better detect the correlation between different frequencies and human activities. The spectrogram demonstrates that different activities exhibit unique frequency energy profiles. By utilizing the energy profile of different frequencies, we can construct an activity recognition model, which quantifies the correlation between the movement speeds of various human body parts and specific human activity. Although the CSI magnitude is clearly related to OOW movement (mainly walking), it remains challenging to identify in-place actions, such as using radar and drinking water. Existing in-place activity recognition techniques are mainly based on machine learning algorithms, which rely on black-box feature exaction and require a labor-intensive and time-consuming process of collecting training data sets, which limits their deployment in ships.

III. PRELIMINARIES & SYSTEM OVERVIEW

In this section, we first introduce the basic MUSIC algorithm and then raise two main challenges in a mobile ship; that is, the low SNR environment and vibrating background reflections. Based on this, we present an overview of our *Wi-Watch* design.

A. Basic MUSIC Algorithm

Accurately extracting the signal path-propagation parameters (such as AoA and ToF) is the key to using WiFi signals to achieve human-environment interactions. Usually, subspace-based algorithms [e.g., MUSIC, estimation of signal parameters via rotational invariant techniques (ESPRIT) [33], and their variants] are used for estimating the signal path parameters in multiple-antenna devices. The basic insight is that signals propagated from paths of different lengths introduce different phase changes at the receiver. Incident signals from different angles across an array of antennas can also introduce corresponding phase shifts according to the distance between each antenna. The MUSIC algorithm constructs the required spatial covariance matrix based on the data from multiple antennas. Moreover, more antennas can provide more array manifold vectors, thus forming a better spatial covariance matrix for a more detailed analysis of signal and noise subspaces. The WiFi NICs provide the CSI measurement matrix (X), and the MUSIC algorithm can be used to estimate the *steering matrix* $A = [\vec{a}(\theta_1), \cdots, \vec{a}(\theta_L)]$, where $\vec{a}(\theta_k)$ is the steering vector of the k-th path. The AoA and ToF can be easily found from the steering matrix; however, we omit the mathematical details here for brevity but refer to the literature discussion about this idea [27]. The standard MUSIC algorithm is an eigenstructure analysis method that can compute the eigenvectors of XX^H and estimate the

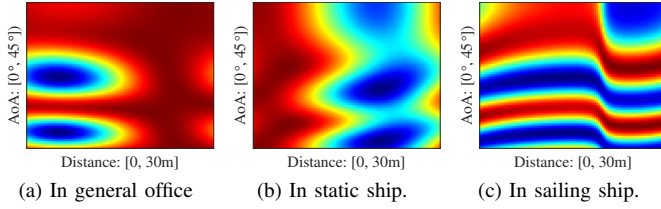(a) In general office   (b) In static ship.   (c) In sailing ship.

Fig. 5: The path parameters (ToF and AoA) estimated by the MUSIC algorithm. The multipath effect in a static ship is more potent than that in a general office environment. When the ship is sailing, the multipath effect will be further aggravated.

Fig. 6: An example of vibration effect on the received signal.

steering vectors orthogonal to these vectors. The correlation matrix of X can be presented as:

$$
\begin{aligned}
R_X &= \mathbb{E}[XX^H] \\
&= A\mathbb{E}[SS^H]A^H + \mathbb{E}[NN^H] \\
&= AR_S A^H + \sigma^2 I,
\end{aligned} \tag{2}
$$

where $S = [s_1, \cdots, s_k]$ is the overall signal, and s_k is the k-th path signal. The n eigenvectors containing the largest eigenvalues correspond to the n signals, and the remainder are noise. The signal and noise subspaces are orthogonal, so the spatial spectrum function is expressed as

$$
P(\theta)_{MUSIC} = \frac{1}{a^H(\theta)E_N E_N^H a(\theta)}, \tag{3}
$$

where $E_N = [e_1, \cdots, e_{M-n}]$ is the noise vector space. Sharp peaks occurred at the AoAs of the incident signals. Although the MUSIC algorithm is promising, it must still overcome two challenges in ship environments.

B. Practical Challenges

1) Low SNR Environment: Subspace-based algorithms mainly rely on the assumption that a strong line-of-sight (LOS) path exists and that multipath signals can be easily decorrelated. However, owing to the *metal structure* and *mobility* of the ship, much environmental noise is generated during its voyages. For the 2.4 GHz WiFi signals, the partition loss of the transmitted power can reach up to 40 dB, whereas most of the radio waves are reflected by the metal surface. The dielectric property of a metal surface can lead to numerous strongly reflected signals and significantly increase the multipath effect at the end of the receiver.

Fig. 5 compares the performances of MUSIC-based path-parameter estimation in a real-world ship environment and general office scenario. Fig. 5b clearly shows that when multi-path echoes are strongly correlated, the AoA-ToF estimation is severely degraded. When the SNR decreases, the orthogonality degrades, which widens and shrinks the peak [34][35]. In this example, the true ground position may not be at the maximum point, and it is challenging to locate. Under the condition of a SNR of -10 dB and four antennas, the calculated value of the Cramer-Rao bound (CRB) for the MUSIC algorithm is 1 degree [36]. This means that the minimum error in AoA estimation is 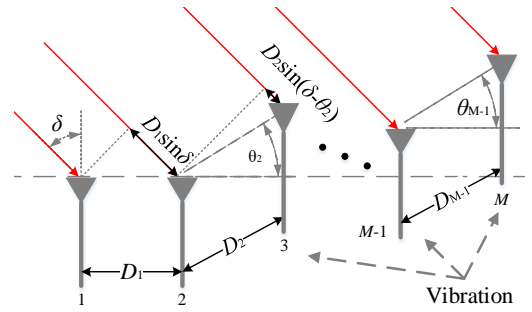1 degree. The dynamic environment of ships presents a significantly low SNR characteristic, which seriously reduces the accuracy of the MUSIC algorithm on ships.

2) Vibrating Background Reflections: There are several complex instruments on the bridge, including a marine radar, an electronic chart system, and a rudder, and they are mainly made of metal, which can significantly affect wireless signals. In addition, when the ship sails, the hull and instruments inevitably vibrate, which introduces dynamic multipath effects. As described above, a multipath signal showing [37] a slightly different signal phase arriving at a receiver would lead a huge variation in magnitude. Interference happens when two path signals superpose to form one resultant signal. The amplitude of the resultant signal may be either larger or smaller than those of the two participating signals, depending on the phase difference between them. The phase offset induced by vibration signals is as follows:

$$
\Delta\phi = \frac{2\pi}{\lambda} \cdot 2\Delta d \mod 2\pi, \tag{4}
$$

where λ is the wavelength. The phase change ($\Delta\phi$) is related to the change in the propagation distance, Δd. According to a previous study, the vibration can be smaller than 100 μm. Owing to the WiFi bandwidth, it is challenging to identify only the signal paths reflected by people out of all the paths.

The problems are as follows. First, only one CSI sample is obtained for each WiFi packet; therefore, the sample interval depends on when the packets arrive. Commodity WiFi devices cannot send and receive packets at fixed precise intervals owing to packet loss/delays caused by environmental noise and interference. Therefore, adjacent CSI samples usually are separated by varying intervals. Second, the WiFi signal is transmitted to multiple subcarriers. Owing to frequency-selective fading, different subcarriers have different SNRs. If we select a subcarrier with a low SNR for processing, the estimated Doppler information will be less accurate. Separating signals of nearby paths is challenging because reflected signals are much weaker than direct-path signals. Therefore, it is difficult to accurately estimate their parameters in interference originating from a strong direct path. Although promising, the MUSIC algorithm alone is insufficient to accurately track OOW vigilance.
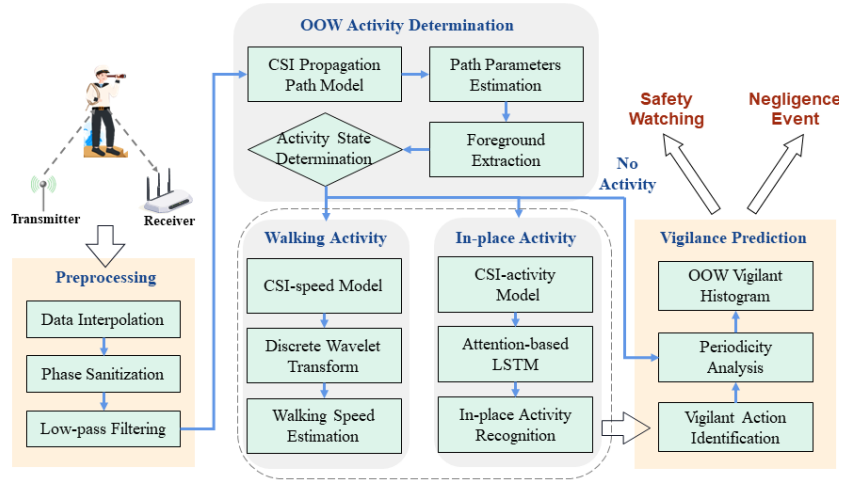
Fig. 7: System overview.

C. System Overview

We propose a privacy-preserving WiFi-based OOW vigilance tracking system *Wi-Watch* for a mobile ship environment. The basic idea of our system is to track the OOW's vigilant actions using WiFi signals by extracting fine-grained signal-path parameters (i.e., ToF and AoA) and evaluating their vigilance according to the OOW's duty rules. The design of the Wi-Watch system includes four modules, i.e., *OOW activity state determination*, *walking velocity estimation*, *in-place activity recognition*, and *vigilance prediction*. To overcome the low SNR and vibrating background, we first adopted the CSI-path model and the MUSIC-space-alternating generalized expectation (SAGE) combined algorithm to accurately estimate the signal-propagation path parameters [i.e., AoA, ToF, and ToF change rate] in a low-SNR environment. Through a novel path-separation algorithm, we can resolve multipath at a much finer resolution, thereby isolating only the signals reflected off targets of interest. Meanwhile, a vibration-background canceling method is proposed to extract the foreground information. An activity state determination module based on the selected human reflection signal paths has been proposed to distinguish between three states: walking activity, in-place activity, and no activity.

Usually, a proper lookout during watchkeeping requires the OOW to walk through the bridge periodically. Therefore, walking plays an important role in OOW duties. To recognize walking, we adopt a *CSI velocity* model, which quantifies the correlation between the change in the signal path distance and human walking velocity. By estimating the OOW's walking velocity, we can evaluate the vigilance of the OOW in a coarse grid. For in-place activities only involving the movements of relatively small body parts, WiFi-based activity recognition is more challenging owing to the limited CSI resolution [38]. Moreover, reflections originating from body parts may propagate through different paths in complex indoor environments. To accurately identify the body-part movement signals, we utilize Doppler-AoA segment pairs to derive a *CSI activity* model. Dynamic time warping (DTW) is used as a similarity metric to align the varied-length activity profiles. The in-

place activity can be recognized by calculating the similarity between a CSI trace and a preconstructed activity profile. Finally, the histograms of the OOW activities are compared with the OOW's duty cycle to achieve a highly robust vigilance prediction metric.

IV. WI-WATCH DESIGN

We first describe data preprocessing and activity determination, followed by presenting walking velocity estimation and in-place activity recognition modules.

A. Data Preprocessing

1) Data interpolation: Wi-Fi packet loss is a common occurrence in real-world settings. To mitigate the potentially negative effects on experimental results caused by sample jitter resulting from such losses, we perform CSI interpolation. We use one-dimensional linear interpolation algorithms to ensure the generation of evenly spaced CSI with a 1-millisecond interval (transmission rate of 1,000 packets/s) between consecutive measurements.

2) CSI cleaning: Due to instances of desynchronization between transmitters and receivers, along with hardware imperfections, the CSI procured by commercial WiFi devices often contains high levels of noise, making it unsuitable for immediate use in channel parameter estimation. To acquire accurate channel estimation results, it is essential to cleanse the CSI data. Thus, we refer to the CSI cleaning approach posited in reference [39]. Primarily, to neutralize nonlinear errors in the CSI, we employ a coaxial cable of a predetermined length to bridge the transmitter and receiver, thereby generating a CSI calibration template. Next, we eliminate the random CSI phase offset caused by the sampling frequency offset and packet detection delay by taking the conjugate multiplication of CSI data from two adjoining antennas on the same Wi-Fi card. Finally, we also remove the strong direct path signal to increase the accuracy of the Doppler estimation of the weak reflection signal.

3) Low-pass filtering: Normal human activities involve limb movements with speeds not exceeding 3 m/s (typical average walking speed ranges from 0.8 to 2.2 m/s, and the swinging speed of arms and lower legs does not exceed 2.5 m/s) [40][41]. The maximum Doppler frequency shift caused by normal human activities in a 2.4 GHz WiFi signal is 48 Hz [42]. Therefore, excessively high-frequency components in CSI are unlikely to be caused by human activities. To mitigate high-frequency noise, we employ Butterworth low-pass filters, which are widely used in signal processing to attenuate high-frequency components while retaining the low-frequency components of a signal. Butterworth filters are known for their symmetric response and are effective in envelope analysis due to their ability to preserve the signal duration. Specifically, we have set the cutoff frequency of the low-pass filter to 200 Hz in order to effectively eliminate high-frequency components in the CSI data that are not attributable to human activities.

B. On-duty Activity State Determination

In this section, we first introduce the **CSI-path Model**. Next, we introduce the **MUSIC-based Joint Parameter Estimation** algorithm. The **Foreground Extraction** module is then presented, which is responsible for extracting the signal reflected from the human body based on the estimated signal path parameters. Finally, we present the **Activity State Determination** module that recognizes the activity state of OOW and measures the duration of each respective state.

1) CSI-path Model: If a WiFi signal propagates through N paths to arrive at the receiver, the CSI measurements for paths at frequency f and sensor (antenna) s can be denoted by the following equation:

$$H(f,t,s) = \sum_{k=1}^{N} \alpha_k(f,t,s)e^{-j2\pi f\tau_k(f,t,s)} + N(f,t,s), \quad (5)$$

where $\alpha_k(f,t,s)$ is the complex value representing the attenuation and phase of the k-th path. $N(f,t,s)$ is the complex white Gaussian noise that captures the background noise. $e^{-j2\pi f\tau_k(f,t,s)}$ is the phase shift caused by the signal path delay (τ_k) and wave frequency f. Owing to the multiple frequency and antenna measurements, the phase shift can be further denoted as:

$$
\begin{cases}
H(f + \Delta f, t, s) = \sum_{k=1}^{N} A_k e^{-j2\pi(f+\Delta f)\tau_k}, \\
H(f, t + \Delta t, s) = \sum_{k=1}^{N} A_k e^{-j2\pi f(\tau_k - \frac{v_k}{c}\cdot\Delta t)}, \quad (6) \\
H(f, t, s + \Delta s) = \sum_{k=1}^{N} A_k e^{-j2\pi f(\tau_k + \Delta s\cdot\phi_k)},
\end{cases}
$$

where Δf, Δt, and Δs are the differences in the subcarrier frequency, packet arrival interval, and antenna distance, respectively. A_k is the abbreviation of $\alpha_k(f,t,s)$; c and v_k denote the velocity of light and the path-length changing rate, respectively. $f_D = f \cdot v_k/c$ is the Doppler frequency shift between the transmitter and receiver, and ϕ_k is the AoA. We can denote the k-th path signal parameters as

$\theta_k = (\tau_k, \phi_k, f_{Dk})$. Next, our goal is to accurately estimate these parameters for all the paths.

2) MUSIC-based Joint Parameter Estimation: To solve Equation 6, we propose a 2D MUSIC algorithm to transform raw WiFi CSI into 2D WiFi AoA-Doppler parameters. Assume that we have received M CSI samples from S antennas in each time window. Considering the CSI window measured from all the antenna arrays, the phase difference between the M CSI samples and the first sample measured from the first antenna can be expressed as:

$$
\begin{aligned}
\vec{a}(v) &= [1, e^{-j2\pi f\frac{v}{c}\cdot\Delta t_1}, e^{-j2\pi f\frac{v}{c}\cdot\Delta t_2}, \cdots, e^{-j2\pi f\frac{v}{c}\cdot\Delta t_M}] \\
&= [1, \quad \Omega_1, \quad \Omega_2, \quad \cdots, \quad \Omega_M],
\end{aligned}
$$
(7)

The phase difference between the S antenna array and the first antenna in one CSI sample can also be expressed as:

$$
\begin{aligned}
\vec{a}(\phi) &= [1, e^{-j2\pi f\Delta s_1\cdot\phi}, e^{-j2\pi f\Delta s_2\cdot\phi}, \cdots, e^{-j2\pi f\Delta s_S\cdot\phi}] \\
&= [1, \quad \Phi_1, \quad \Phi_2, \quad \cdots, \quad \Phi_M],
\end{aligned}
$$
(8)

Steering vectors $\vec{a}(v)$ and $\vec{a}(\phi)$ are caused by the movement of the Doppler velocity and the signal path arriving at AoA. Therefore, the received CSI sample matrix can be represented as:

$$
\begin{aligned}
\mathbf{H}(f) &= \begin{bmatrix} H(t_0, s_0) & \cdots & H(t_0, s_0 + \Delta s_S) \\ \vdots & \ddots & \vdots \\ H(t_0 + \Delta t_M, s_0) & \cdots & H(t_0 + \Delta t_M, s_0 + \Delta s_S) \end{bmatrix} \\
&= \begin{bmatrix} 1 & \cdots & \Phi_S \\ \vdots & \ddots & \vdots \\ \Omega_M & \cdots & \Omega_M \cdot \Phi_S \end{bmatrix} \times H(f, t_0, s_0) + N(f) \\
&= AS(f) + N(f),
\end{aligned}
$$
(9)

The 2D MUSIC algorithm can now be applied to jointly estimate the AoA and Doppler shift of each path using Equation 3.

3) Foreground Extraction: The path parameters estimated by 2D-MUSIC algorithm consist of multiple path signals, including reflection signals from OOW and vibrating objects within the bridge. As a result, a filtering algorithm is necessary to extract the reflection path parameters generated by human movements and remove the vibration noise. The key observation is that the change in path parameters caused by human movement is more continuous and regular compared to the change caused by the vibration of objects inside the vessel, which is random. In this section, we apply the graph-based path matching (GPM) algorithm proposed in Widar2.0 [43] to extract the reflection path parameters of human movement, namely foreground, while eliminating the reflection path caused by the vibrating objects. The GPM algorithm combines the two sets of parameters with the smallest difference between adjacent moments and identifies the combination of parameters with the highest weight as the reflection path parameters of continuous human movement.

The principle of the GPM as shown in Fig. 8, suppose that T time-series CSI packets are measured for K signal paths in the environment. θ_k denotes that estimated signal parameters
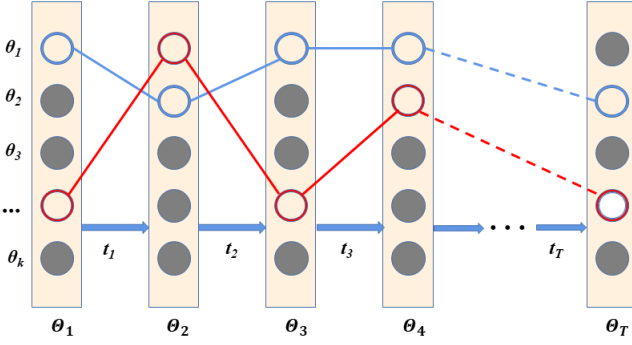
Fig. 8: The principle of GPM.

of the k-th path. Θ_t denotes that estimated signal parameters in the t-th packet. θ_{kt} denotes the estimated signal parameters of the k-th path in the t-th packet. Next, a weighted T-partite graph G=(θ_{kt}, E, W) is defined, where $e_{k_1t_1}^{k_2t_2} \in E$ represents the edge between $\theta_{k_1t_1}$ and $\theta_{k_2t_2}$ at any adjacent moments; and $w_{i_1j_1}^{i_2j_2} \in W$ represents the weight of the edge $e_{k_1t_1}^{k_2t_2}$. The weight is defined as the distance between parameters:

$$w_{k_1t_1}^{k_2t_2} = \|c\left(\theta_{k_1t_1} - \theta_{k_2t_2}\right)\|, \tag{10}$$

where c is the vector of coefficients that normalize different parameters AoA and DFS. We then denote $x_{k_2t_2}^{k_1t_1}$ as binary variable that indicates whether the edge $e_{k_1t_1}^{k_2t_2}$ is selected for matching. Thus, the objective function is:

$$x_{opt} = \arg\min_{\mathbf{x}} \mathbf{w}^T\mathbf{x}, \tag{11}$$

where w and x are vectorized weights and binary variables respectively. To ensure that selected edges form K T-order complete graphs, several constraints should be fulfilled.

$$\begin{cases} x_{kt_1}^{kt_2} = 0, \\ \sum_{t_2=1}^{T} x_{k_1t_1}^{k_2t_2} \leq 1, \\ \sum_{t_1=1}^{T}\sum_{k_2=1}^{K}\sum_{t_2=1}^{T} x_{k_1t_1}^{k_2t_2} = T \cdot K - T, \\ x_{k_1t_1}^{k_2t_2} + x_{k_1t_1}^{k_3t_3} \leq 1 + x_{k_1t_1}^{k_3t_3}, \end{cases} \tag{12}$$

This optimization problem can be solved using a binary integer program (BIP). By optimizing the objective function, the two sets of parameters with the smallest difference between two adjacent moments are found and combined into a reflection path. Because the signal parameters reflected from people change slowly with time, the background reflection parameters change significantly at each time step. Equation 11 can be used to eliminate the noisy reflection and determine the OOW's reflected signal. As shown in Fig. 8, the blue and red plots indicate the path parameters reflected by people and the environment, respectively. In addition, the static paths are removed by subtracting the mean value from the conjugate multiplication to avoid interference with the OOW's reflecting signal selection because of its high power before path selection.

4) Activity State Determination: This module is designed to detect the activity states of OOW, along with their respective start and end times. We propose that OOW activity states fall into three categories: walking, in-place, and no activity. After obtaining the foreground parameters from the CSI measurements, Wi-Watch then segments the AoA-Doppler time series into several traces and determines the activity state by analyzing the changing AoA-Doppler pattern. Specifically, we first segment the AoA-Doppler time series, with a segment interval of 5 seconds. Then, the AoA-Doppler sequence segments are matched with the trained HMM models of different activity states to obtain posterior probabilities for determining the OOW's activity state. Finally, the OOW activity state and the start and end time of the activity state are determined. When using HMMs to determine activity, it is assumed that the observed feature vector sequence, corresponding to an activity, is generated by a Markov model. The Markov model is a finite state machine that changes state once per time unit. Each time a state is entered, a feature vector is generated from an output probability density also known as the emission probability distribution. Additionally, the transition between states or the state loop has a probability and is dictated by a discrete probability distribution called the transition probability distribution. HMM models have the capability to capture information from all training samples and are therefore effective, even when there is high within-class variance. For further information regarding HMM model training and classification, please refer to Section V.

C. Walking Velocity Estimation

We propose a CSI-velocity model to estimate the moving velocity of the OOW for estimating vigilance. The duration and moving velocity are the key evaluation criteria for an OOW's lookout. In this module, we first introduce the CSI-velocity model. Next, we introduce the walking speed estimation method based on wavelet transformation.

1) CSI-Velocity Model: Although we can obtain the Doppler shift of each path reflected from the human body using the previous method, the extracted Doppler frequency shifts are not the real velocity of the moving OOW. Owing to the complex shape of human bodies, wireless signals can be reflected from different body parts moving at different velocities. All the body-part velocities jointly determine the Doppler shifts of the target-reflecting paths. An advanced CSI-velocity model that outputs movement information is required to estimate the real walking velocity of OOWs from multipath Doppler shifts. According to CSI, the Doppler frequency shift of the signals extracted from reflected off a moving human body can be represented by

$$|H(f,t)| = |H_s(f,t)| + \sum_{k \in N_d} |\alpha_k(f,t)| \cos(2\pi f_{Dk} + \phi_{sk}), \tag{13}$$

where the static component $H_s(f)$ is a constant vector, while the dynamic component is a linear combination of a set of reflected paths represented by sinusoids. The frequency of the sinusoids is related to the Doppler shifts. Therefore, the

OOW's walking velocity can then be estimated by extracting each frequency component using time-frequency analysis tools.

2) Walking Velocity Estimation: To estimate the OOW walking velocity, we used a discrete wavelet transform (DWT) to separate these components in the frequency domain. According to IEEE 802.11, at a frequency of 2.472 GHz, the subcarrier wavelength is approximately 12.14 cm. In our observation, the frequency of continuous OOW walking is mainly in the range 15-20 Hz, while the frequency of in-place activities is mostly below 10 Hz. The continuous-walking-induced Doppler shift can be obtained by measuring the frequency, and the corresponding velocity of the path-length change is approximately 1.8-2.4 m/s. Considering that the wireless signal travels back and forth, the velocity of the moving human body is approximately half of the change in the signal-path velocity, which is approximately 0.9-1.2 m/s. To further extract the frequency component, wavelet decomposition was performed on the filtered CSI power. Decomposing one level can obtain the approximation coefficient (CA) and detail coefficient (CD), which represent low- and high-frequency information, respectively. Using the Harr wavelet function, the CSI signal is processed using four levels of wavelet decomposition. The first level obtains the coefficient (CD_1) of the high-frequency band (55-110 Hz), where CD_1 represents the moving velocity of 3.3-6.7 m/s in the 2.4-GHz frequency band. CD_2 of the second level indicates the range 27.5-55 Hz, and the corresponding walking velocity is 1.7-3.3 m/s. After four-level decomposition, the wavelet coefficient matrix ($X = [CA_4, CD_4, CD_3, CD_2, CD_1]$) indicates the energy in each frequency. We can obtain the OOW's walking velocity by measuring the energy of each frequency band.

D. In-place Activity Identification

Once the activity states discrimination module identifies an in-place activity, it segments it and forwards it to the in-place activity recognition module. OOW activity durations generally vary; for instance, the time required for a crew to visually survey an area using binoculars may range from 3 to 6 seconds, depending on the situation. Consequently, all OOW activities were normalized to a standard length with dynamic Time Warping (DTW). Additionally, considering that the AoA-Doppler sequence is a time series data, we used the LSTM model to recognize the OOW in-place activities.

1) CSI-activity Model: The proposed attention-based LSTM framework is shown in Fig 9. First, a sliding window consisting of Doppler-AoA pairs is fed as input into an LSTM network for automatic OOW-activity feature learning. An LSTM cell contains a hidden state (h), a cell state (c), and three gates: a forget gate (f), an input gate (i), and an output gate (o). In every step, input $H_g^{(t)}$ is stacked with the previous hidden state (h_{t-1}) and is multiplied by LSTM weights W_L to determine the values of all the gates and the variable (g). The three gates are used to determine the current inner states (h_t) and (c_t). Among them, f decides whether to forget the previous state (c_{t-1}), i decides whether to read the new input ($H_g^{(t+1)}$) and determines the current cell state (c_t) with g.
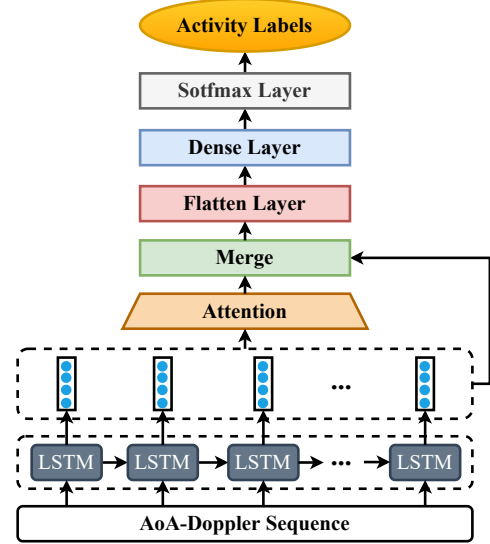

Fig. 9: The proposed attention-based LSTM framework for CSI-activity model.

Finally, the current state (h_t) is calculated as o and c_t. The LSTM cell update is defined as follows:

$$\begin{pmatrix} i \\ f \\ o \\ g \end{pmatrix} = \begin{pmatrix} \sigma_L \\ \sigma_L \\ \sigma_L \\ \tanh \end{pmatrix} W_L \begin{pmatrix} h_{t-1} \\ H_g^{(t)} \end{pmatrix}, \tag{14}$$

$$c_t = f \odot c_{t-1} + i \odot g, \tag{15}$$

$$h_t = o \odot \tanh(c_t), \tag{16}$$

where, $\odot$, σ_L, and $\tanh$ represent the element-wise multiplication, logistic sigmoid activation, and hyperbolic tangent functions, respectively. W_L is the weight of the LSTM cell, which is used for the AoA-Doppler sequence transfer. A Tanh-based score function was used to evaluate the importance of each feature. The attention model is designed as a layer to normalize each feature at every time step and indicates the importance of the features. The score (s_i) of the attention is obtained as follows

$$s_i = \tanh(\mathbf{W}^T h_i + b), \tag{17}$$

where $\mathbf{W}^T$ and b represent the weight and bias, respectively. We obtain the attention matrix by utilizing the hidden states of LSTM through element-wise multiplication with the LSTM output. Then, the feature matrix is transformed into a feature vector by a Flatten Layer. Three dense layers are used to map high dimensional feature vectors to low dimensions. Finally, a softmax classification layer is used to identify different 6 in-place activities with the final feature vector.

2) Model Training: The proposed attention-based LSTM framework is trained to determine all the model parameters based on training data with true labels. Initially, all parameters are randomly assigned, and then the training data is inputted into the attention-based LSTM to predict the labels. Each activity sample is normalized to the same length of 10 seconds. It is split into 50 groups of vectors with an interval of 0.2 seconds

TABLE I: The parameters of proposed attention-based LSTM model.

Parameters for deep learning model	Values
The units in LSTM	400
The number of attention hidden state	128
Learning rate	1e-4
Batch size	128
Epochs	60

and sequentially fed into the LSTM model. Each group of vectors has a length of $200 \times 2 = 400$. The LSTM model has a total of 128 hidden state units. The predicted labels and the given true labels are utilized to calculate the categorical cross-entropy loss. Gradient-based optimization methods are applied to update the model parameters through back-propagation. We use adaptive moment estimation (ADAM) for this purpose, as it can effectively compute adaptive learning rates for the parameters during optimization. The finalized parameters of the attention-based LSTM model are demonstrated in Table I

E. Vigilance Evaluation

We have defined the low vigilance state of OOW during on-duty hours, and Wi-Watch system must sound an alarm immediately upon detecting this state. The low vigilance state includes off-duty, sleep duty, and prolonged inactivity. According to the Regulations on Watching Crews at Sea, the crew on duty on the bridge must be composed in such a manner that there is no chance of anyone being left unattended under any circumstances. In addition, the bridge prevents the driver from sleeping at any time. If the driver remains inactive for more than 3 minutes, as indicated by the test result, an early warning will be triggered. It is imperative that the driver continuously performs tasks such as measuring and plotting charts and other related operations during the duty period, while also being aware of the surrounding navigation information.

V. EXPERIMENTS & EVALUATIONS

We fully implemented and extensively evaluated our proposed system on an actual ship. We first introduce the testbed and experimental setup. After that, the system evaluation and performance are described.

A. System Implementation

Wi-Watch consists of one transmitter and at least one receiver. The transmitter employs a TP-LINK router, which operates on the 2.4 GHz frequency band and has a 20 MHz bandwidth. The receiver employed a standard ThinkPad T400 laptop, outfitted with a commercial off-the-shelf 802.11n 5300 NIC and the operating system Linux kernel 2.6.34. The model training and testing are performed by a Linux desktop with a 3.40 GHz Intel® E3-1231 CPU, 15.6 GB of RAM, and Nvidia GTX 1080 GPU. The Wi-Watch components were implemented using Python. For the deep neural network, we constructed the attention-based LSTM network using Pytorch. All the experiments that we report in this paper were performed in
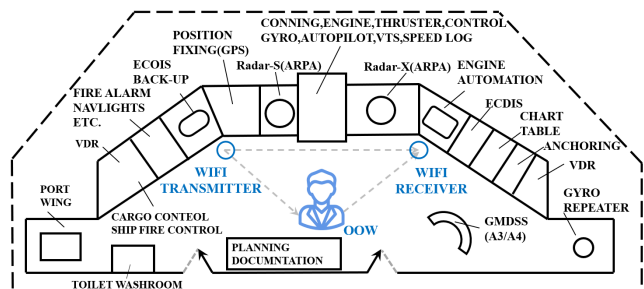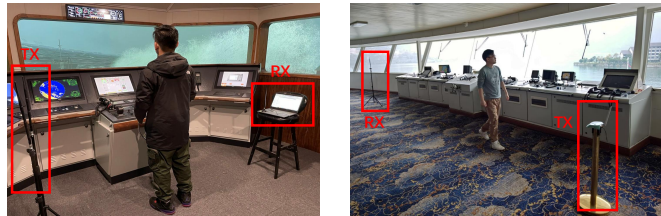

Fig. 10: A typical setup of devices and environment.

(a) In simulator.

(b) In ship.

Fig. 11: Experiment scenarios of a ship driving simulator and a real-world passenger ship.

the 2.4 GHz frequency band with 20 MHz bandwidth channels. Note that the 802.11n CSI tool only provides CSI values of 30 sub-carriers even though a 20 MHz WiFi channel has 52 sub-carriers. We implemented our experimental testbed on a real-world cruise ship dubbed "Yangtze Goddess II" from Chongqing to Yichang. The WiFi transmitter and receiver were placed 3 m apart under line-of-sight (LOS) conditions. The transmitter activates one antenna and broadcasts Wi-Fi packets at a rate of 1,000 packets per second. We have set the transmit power to a constant 15 dBm, which is a common transmit power for WiFi devices operating in the 2.4 GHz band and complies with China's regulations on Effective Isotropic Radiated Power (EIRP) [44]. Considering the presence of an automatic gain control (AGC) module in WiFi, it is necessary only to incorporate an AGC error elimination module [39]. The receiver activates all three antennas which are placed in a line. As shown in Fig. 10, the user is not on the direct path between the transmitter and receiver. This is a normal case in real life.

B. Evaluation setup

To accurately evaluate the Wi-Watch performance, the experiment was carried out on an actual cruise ship ("Goddess II") from Chongqing to Yichang and in a ship-driving simulator at the Wuhan University of Technology. Fig. 11 shows the interior of the simulator and the real-world ship bridge, where the transceivers were placed 1 m above the floor. The ship bridge was on the 5th floor in a 19m × 6m room, which is the key area where OOWs perform their duties. A ship ship-driving simulator, which is designed to mimic the instrumentation and facilities of real-world ships, was employed in a 6m × 6m room to train the crew's skills. To label the raw CSI data, 7 OOW activities were collected

TABLE II: Information about the 15 experimental assistants.

Name	Gender	Age	Ht(cm)	Wt(kg)
A	F	25	163.1	56.2
B	F	28	165.3	58.5
C	F	30	168.5	82.5
D	M	23	172.4	85.1
E	M	25	180.7	73.6
F	M	24	172.5	62.6
G	F	23	165.8	46.9
H	M	24	170.8	68.4
I	M	31	176.6	80.1
J	F	21	168.6	57.6
K	F	21	187.2	76.5
L	M	27	180.2	78.4
M	M	26	174.1	64.2
N	F	24	173.7	61.6
O	M	26	178.7	85.6

using a camera, including walking, reading, steering, drinking, using the radar, using binoculars, and using a smartphone. We assume that only the experimental assistant is present in the sensing area (between the transmitter and the receiver), as moving entities would introduce noise from reflected signals, thereby leading to a less accurate estimation of the signal path parameters for the target assistant. Each activity was performed 20 times for 60 s by 15 experimental assistants (8 males and 7 females). At the end of the activity, the assistant remained still. The weight of the assistants varies from 56.2 kg to 85.6 kg, the age varies from 21 to 30 and the height varies from 163.1 cm to 187.2 cm. And the details of the experimental assistant information are illustrated in Table II.

C. Data Description

We collected data from two scenarios. Our dataset consists of three driver activity states, including seven typical on-duty activities within the in-place activity state. Prior to data collection, we required the experimental assistant to watch example videos and sketches of each activity. The sketches of the OOW activities are plotted in Fig. 3. Furthermore, the experiment assistant conducts several in-place and walking activities which are not profiled. They are used to evaluate the robustness of our system for recognizing unknown or random activities. Therefore, there are over 3,600 activity samples (15user×20times×6activities×2scenarios) in total. We divide the dataset into training and testing randomly with a ratio of 9:1. Note that the data are collected on different days.

D. Performance Metrics

We use both the confusion matrix and recognition accuracy to evaluate the performance of our system.

Confusion Matrix. Each row represents the actual activity performed by the experimental assistant and each column shows the activity it was classified as by Wi-Watch. Each cell in the matrix corresponds to the fraction of activity in the row that was classified as the activity in the column.

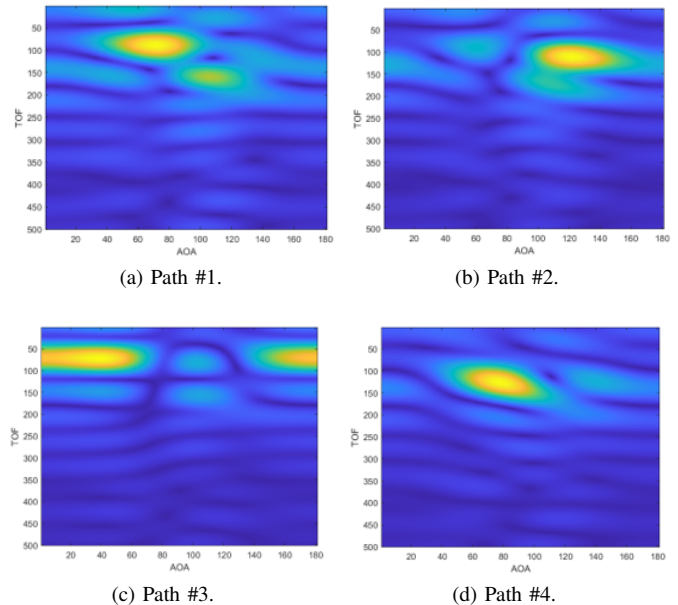**Recognition Accuracy.** The percentage of the activities correctly classified by our system.

(a) Path #1. (b) Path #2. (c) Path #3. (d) Path #4.

Fig. 12: Doppler shift and AoA estimation of four paths from raw CSI.

E. System Performance

1) Baseline Methods: We compared Wi-Watch with 4 state-of-the-art CSI-based activity-recognition programs (CARM, WiSH, and E-eyes, Mosense) in a mobile ship environment. CARM [45] is the Baseline method that uses off-the-shelf WiFi NICs and correlates CSI power dynamics with human motion to determine activities such as running, walking, and falling. WiSH [46] is a real-time human presence detection system that can shift CSI eruptions due to irrelevant instantaneous motions or environmental changes. E-eyes [47] compares the CSI measurements with the location-activity profile to recognize daily activities associated with several specific locations in the home. Mosense [48] designs a distance-based subcarrier selection algorithm that captures the different impacts of human movements on WiFi signals by capturing the similarities between subcarriers.

2) Doppler-AoA Estimation: We show the performance of the proposed MUSIC-based Doppler-AoA estimation algorithm. Fig. 12 plots the results obtained when an experimental assistant walked on the ship bridge at 1.5 m/s. The x- and y-axes represent the path AoA and ToF parameters estimated for each path, respectively. The ToF resolution is 0.2 × 50 ns = 10 ns and AoA is $1°$. Because the ToFs and AoAs of the hand and body reflections were different in each estimation, the different propagation paths had to be separated, and the parameters had to be estimated individually. Clearly, our proposed method separated the four main paths easily, and the peaks merged into a 2D profile at each path. Paths #2 and #4 clearly show the main reflection object, the moving OOW, in the area. The distance and AoA estimation of the two paths were differentiable in the 2D profiles. However, the multipath impact is severe on ship bridges, especially the reflectors that vibrate while the ship is sailing. When reflectors are vibrating on the
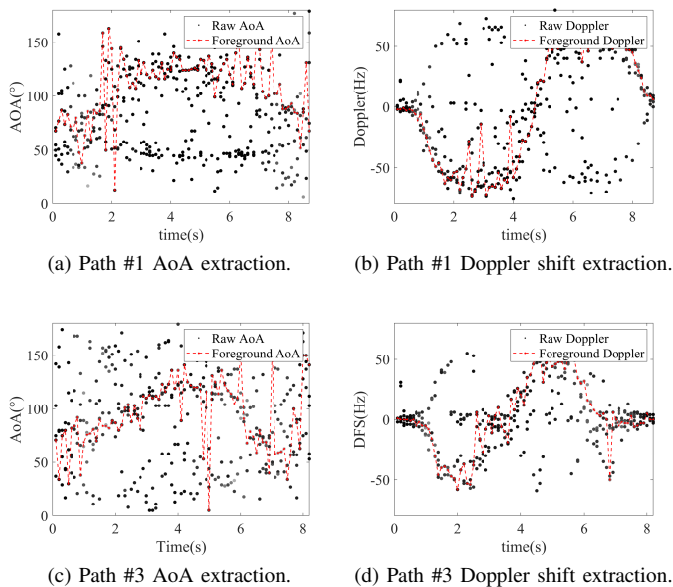
(a) Path #1 AoA extraction.

(b) Path #1 Doppler shift extraction.

(c) Path #3 AoA extraction.

(d) Path #3 Doppler shift extraction.

Fig. 13: Foreground signal paths extraction for low SNR ship environment.

TABLE III: The time overhead of all samples in the activity state determination module during the training and testing stages.

Activity state	Training size	Training time(s)	Testing size	Testing time(s)
Walking	332	39.42	30	0.42
In-place	288	35.48	30	0.48
No activity	190	25.65	30	0.40

bridge, the orthogonality of the MUSIC matrix may degrade, possibly resulting in multiple peaks, as shown in paths #1 and #3. Therefore, the foreground parameters (which are the signal paths reflected from the OOW) must be identified, and the signal paths must be eliminated from the irrelevant objects before determining the OOW's activities. Practical tests revealed that our proposed AoA-Doppler parameter estimation algorithm, based on the 2D-MUSIC, can estimate a signal with a length of 0.1 seconds (sampling rate 1000 packets/s) in 0.001 seconds. For a signal with a length of 10 seconds, we used a time window and a sliding window, both of 0.1 seconds in length, allowing the algorithm to complete in 34.088 seconds.

3) Foreground Extraction: To evaluate the performance of the proposed foreground extraction method, we collected several traces without placing an object in our experimental setup. The dark dots in Fig. 13 show the typical multipath signal parameters of a trace at each time step, where the experimental assistant walked from the radar to the steering wheel and then turned back. We also weighted the color of the dots to illustrate the parameters and their corresponding path power. Clearly, the AoA and Doppler-shift parameters mainly exhibited 2 paths from time 0 to 8 s. Although the starting and ending points of both paths were similar, the changing patterns were opposite. Although both inverse-changing path parameters can be distinguished by visual inspection, it is difficult to classify the parameter dots because the multipath parameters are confusing together. Moreover, the estimated parameters are mixed with surrounding environmental noise and the multipath originating from the vibrating reflectors. Using our proposed foreground extraction method, the OOW's reflected signal path has been clearly selected out of the cluttered estimations and is plotted in red. Although the ToF estimation is limited by the strong noise in the metal environment and the low resolution of commercial WiFi NICs, the Doppler shift caused by the changing ToF can be used to describe the signal propagation

variation as a fine-grained recognition feature. Therefore, the AoA and Doppler shift changes introduced by walking OOWs can be obtained in such a low-SNR ship environment. Practical tests have demonstrated that our foreground extraction algorithm, which is based on the GPM, can match an AoA-Doppler sequence for a signal path of five seconds in duration within 0.136 seconds.

4) Activity State Determination: To evaluate the performance of the HMM-based activity state determination model, 10 sets of OOW activity state samples, belonging to each of the 3 activity state types, were selected as test samples. These samples were then respectively input into the 3 established HMM models, and the average of the maximum output probabilities obtained by the test samples in the HMM model library was computed. In order to display the calculation process of the maximum probability output of each sample in HMM model, we selected randomly one sample as a representative to show the computation matching process of the test samples in the HMM library. The calculation results are shown in Fig. 14. The output probability of each state sample in the corresponding HMM is the largest. Additionally, we compare the overall performance of our method to four state-of-the-art activity state recognition methods and present the recognition accuracy of each activity state using each method. In this experiment, we collected a dataset of 900 activity state samples in ship environments and utilized 90% for training and 10% for testing purposes. Table II illustrates the time allocations for all test samples during the testing phase. As shown in Fig. 14d, our method achieved an average accuracy of 94.38% for all 3 activity states, which is more than 10% higher compared to other methods, representing a significant improvement. This indicates that our system can remove the noise brought by ship vibrations to WiFi signals and achieve robust activity state recognition.

5) In-place Activity Recognition: In this section, we evaluate Wi-Watch's ability to recognize in-place activities. We randomly selected a total of 2,400 samples from different users, including 6 kinds of in-place activities, and divided them into training sets and test sets with a ratio of 9:1. The results indicated that the SVM machine-learning algorithm employing handcrafted features exhibited the poorest performance. The HMM model exhibited slightly better performance compared to the SVM-based approach. In contrast, the LSTM model demonstrated superior performance when compared to both the HMM and SVM models that employed handcrafted features, effectively extracting features automatically. Moreover, the LSTM network also incorporated temporal dependencies in sequential data. Leveraging the proposed attention mechanism and feature-learning techniques, the attention-based LSTM
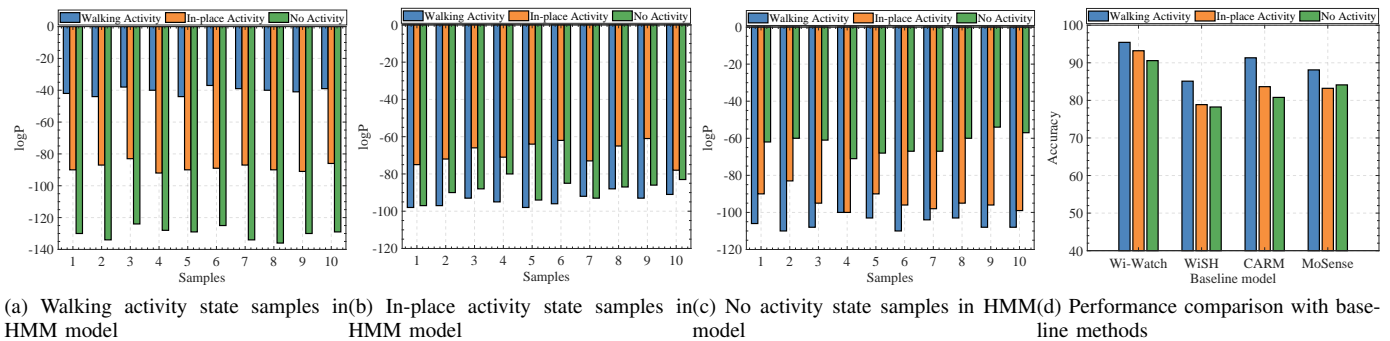
(a) Walking activity state samples in HMM model

(b) In-place activity state samples in HMM model

(c) No activity state samples in HMM model

(d) Performance comparison with baseline methods

Fig. 14: Matching results of the test sample in HMM model.

(a) SVM
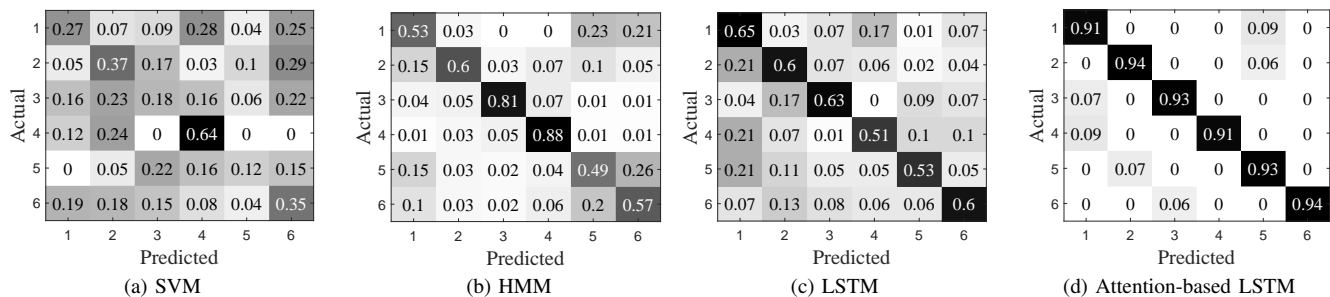
(b) HMM

(c) LSTM

(d) Attention-based LSTM

Fig. 15: The confusion matrix of in-place activity recognition with four methods. The labels numbered 1 through 6 correspond to the following actions: using radar, using binoculars, using smartphones, reading, driving, and drinking.

TABLE IV: The time overhead of the four activity recognition methods during the training and testing stages.

Time(s)	SVM	HMM	LSTM	Attention-based LSTM
Training (s) (2160 samples)	144.12	230.56	5168.86	8012.60
Testing (s) (340 samples)	1.72	2.85	4.39	6.22

(a) In-place activity recognition.

(b) Activity state recognition.

Fig. 16: Impact of ship sailing speed.

model achieved the highest performance in accurately recognizing all six OOW activities. The activities were assigned labels 1-6, respectively, as follows: "Using radar," "Using binoculars," "Using smartphone," "Reading," "Steering," and "Drinking." The average accuracy of the four aforementioned methods in the in-place activity recognition task was compared. The confusion matrix depicting the results of the four methods is presented in Fig. 15. Notably, the attention-based LSTM model attained an impressive overall average accuracy of 94%. Table IV represents the time expenditures on the training and testing phases for both the attention-based LSTM model used in our study and the alternative models explored for comparison.

F. Efficiency and Scalability

1) Impact of ship sailing speed: Given that a moving ship environment differs significantly from stationary land environments, the motion and dynamics of a ship may impact the system's recognition ability. An increase in ship sailing speed
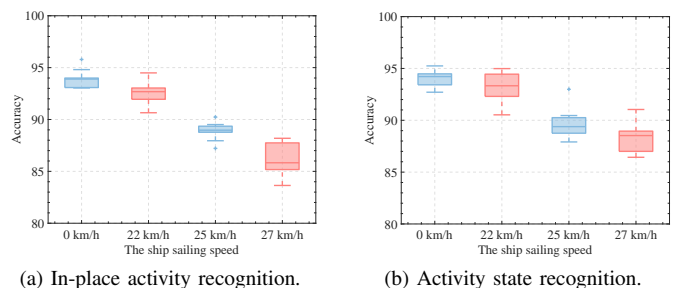
leads to more intense ship vibration, introducing more noise to the WiFi signal. To evaluate the robustness of Wi-Watch at different ship sailing speeds, we collected CSI datasets at different ship speeds and used 5-fold cross-validation to test the average accuracy of Wi-Watch in in-place activity recognition and activity state recognition, as shown in Fig. 16. We use a ship-mounted GPS satellite navigator to obtain the current sailing speed of the ship. The experimental results revealed that the system achieved the highest accuracy in both in-place activity recognition and activity state recognition, surpassing 95.32% when the ship was anchored. With an increase in a ship sailing average speed from 22 km/h to 27 km/h, the average accuracy rate of Wi-Watch's 6 OOW in-place activity recognition decreased from 94.2% to 86.1%, while the average accuracy rate of activity state recognition decreased from 94.15% to 88.14%. The most significant decrease was
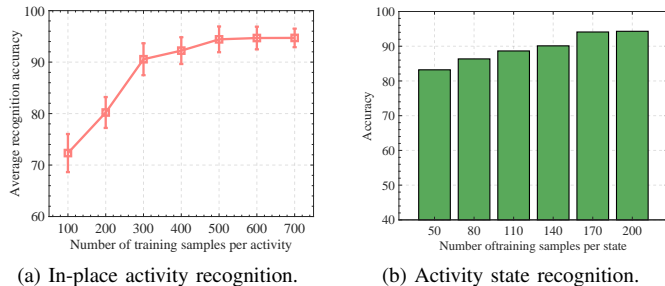
(a) In-place activity recognition.　　(b) Activity state recognition.

Fig. 17: Impact of training set size.

(a) In-place activity recognition.　　(b) Activity state recognition

Fig. 19: Impact of training diversity.

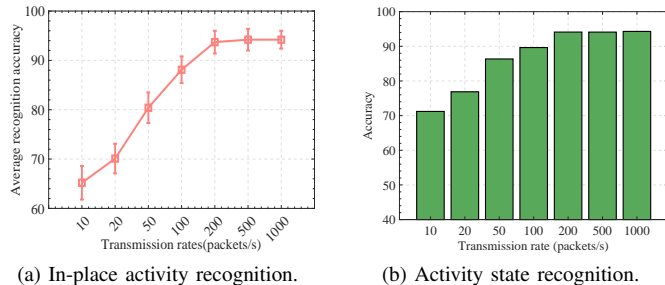(a) In-place activity recognition.　　(b) Activity state recognition.

Fig. 18: Impact of the transmission rate.

observed in the accuracy of in-place activity state recognition, which decreased by approximately 12%. This is because the in-place activity involves small body movements, resulting in less variation in the DFS signals captured by WiFi. The recognition of in-place activity DFS features by Wi-Watch is interfered with when the signal noise level increases.

2) Impact of training set size: The performance of the learning-based method is strongly correlated with the size of the training set. It is necessary to study the impact of training size on system performance. Here we define the training size as the number of training samples for each activity. As shown in Fig. 17a, our system can achieve consistently high accuracy on in-place activity recognition with different training sizes. Especially, our system maintains over 94% accuracy on 6 OOW in-place activities recognition even with the training size as small as 300. Additionally, we evaluated the impact of training set size on the HMM-based activity state determination model. We demonstrate the average accuracy of the model in recognizing the three types of activity states under different training set sizes, as shown in Fig. 17b. We gradually increased the number of training set samples from 50 to 200, and the average accuracy of the model in recognizing the three activity states increased from 83.23% to 94.33%.

3) Impact of transmission rates: Another important factor that influences the recognition accuracy of Wi-Watch is the CSI sampling rate. The use of Wi-Watch for In-place activity recognition necessitates packet transmission, which may interfere with normal communication flow. Therefore, we evaluated the performance of Wi-Watch with 5-fold cross-validation at different CSI transmission rates. We collect CSI measurements at the initial transmission rate of 1,000 packets/s and down-sample the CSI series to 500 packets/s, 200 packets/s, 100
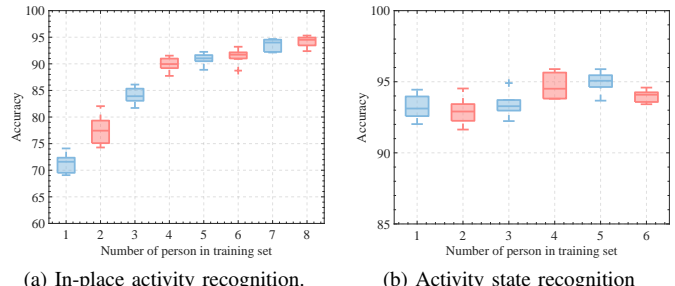
packets/s, 50 packets/s, 20 packets/s, and 10 packets/s. The evaluation results are shown in Fig. 18a. When the sampling rate drops from 1,000 packets/s to 100 packets/s, the average in-place activities recognition accuracy decreases from 94.21% to 88.16%. When the sampling rate reduces to 10 packets/s, the average recognition accuracy decreases to 65.21%. Based on these results, the sampling rate should be no larger than 100 packets/s to get high recognition accuracy. Additionally, we evaluated the impact of transmission rate on the performance of the activity state determination model, and we demonstrate the accuracy of the three types of activity recognition at different transmission rates. As shown in Fig. 18b, when the transmission rate is greater than 200 packets/s, the accuracy of the cross-validation is above 94%. Although the transmission rate gradually increases to 1,000 packets/s, the accuracy of activity state recognition remains relatively constant, and the system performance is no longer affected by the transmission rate. Therefore, it can be concluded that the accuracy of Wi-Watch increases when the CSI is calculated with a higher transmission rate, but the increase is not significant beyond the sampling rate of 200 packets/s.

4) Impact of training diversity: We recognize that the inclusion of diverse experiments assistants in the training dataset can influence the performance and generalization ability of the Wi-Watch system. We vary the number of experimental assistants from 1 to 8, ensuring a range of individuals with different characteristics are included in the training data. We then use data from a new person, who was not part of the training dataset, to test the performance of the Wi-Watch system. Fig. 19a shows that the 6 OOW in-place activities' average recognition accuracy decreases from 94.13% to 72.16% when the number of people for training varies from 1 to 8. This indicates that as the number of training sets increases, Wi-Watch learns more generalized activity features that are independent of specific individuals. Additionally, we evaluated how the diversity of training sets impacts the activity state recognition task, as demonstrated in Fig. 19b. Despite increasing the number of individuals from 1 to 5, the accuracy of activity state recognition did not improve significantly. This is due to the significant differences in the AoA-Doppler features produced by various active states, making it easy for the model to recognize the differences between them.

5) Impact the number of hidden states: Choosing the appropriate number of hidden states during HMM-based OOW

(a) Walking activity of different hidden states. (b) In-place activity of different hidden states. (c) In-place activity of different hidden states. (d) Training of walking activity. (e) Training of in-place activity. (f) Training of no activity.
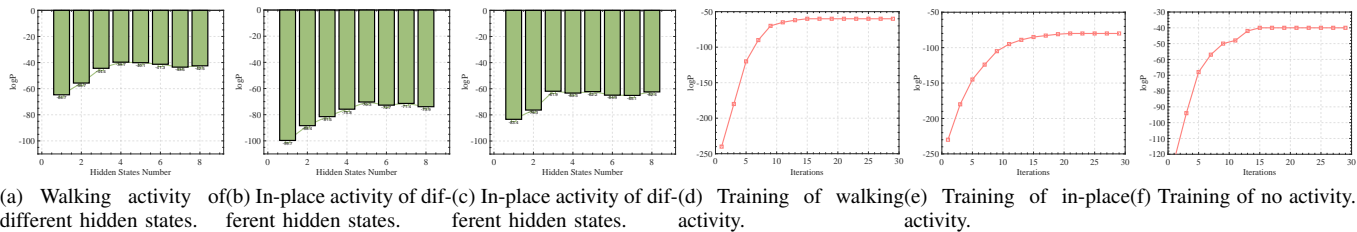
Fig. 20: Performance of the activity state determination model in different parameters.

TABLE V: The score of the model under different numbers of hidden state.

HMM model	3	4	5	6	7	8	9	10
λ_ω	-64.7	-55.7	-44.4	-39.7	-40.1	-41.3	-43.5	-42.5
λ_p	-99.7	-88.4	-81.5	-75.8	-70.3	-72.7	-71.4	-73.9
λ_s	-83.4	-76.3	-61.9	-63.3	-62.2	-64.9	-65.1	-62.4

activity state recognition training is crucial as it directly impacts both performance efficiency and recognition accuracy. The choice of the number of states and the training of each model is performed using the Baum-Welch algorithm and the cross-validation principle: the optimal number of states and the associated optimal HMMs parameters are those yielding the highest likelihood on the validation set. We randomly selected 900 OOW on-duty activity samples representing the three on-duty activity states (walking activity, in-place activity, and no activity) in the dataset (Note that these activity data were randomly selected from different experimental assistants). To increase the accuracy of the recognition model and avoid gaining training costs, we used 3 to 10 state numbers to successively train each on-duty state. We ran 5-fold cross-validation on 900 active samples to ensure that the optimal number of hidden states could be determined. The average evaluation scores of the model posterior probabilities were calculated using different hidden states. Table V shows the average HMM scores obtained for the on-duty activity states (i.e., walking, in-place activity, and remaining still) for different numbers of hidden states. The HMM parameters of the three duty states, state, and no-action states, are represented by λ_w, λ_p, and λ_s, respectively. The higher the score, the more accurate the model recognition under the given number of hidden states. It was evident that, for the six hidden states, the walking state was recognized the most accurately by the model. The in-place and no-activity state models had the optimal number of hidden states set to 7 and 5, respectively. To present the relationship between the model accuracy and the number of hidden states intuitively, histograms for models with different numbers of hidden states are plotted for the three types of activity states, as shown in Fig. 20a 20b 20c. The results show that the local optimal output probability increases with the increase of state numbers, and the model fit becomes increasingly higher. When the local optimal probability no longer increases significantly, the optimal HMM model is determined, with the corresponding number of states being the optimal hidden state number. Fig. 20d 20e 20f shows the specific training iteration process for each type of activity state to obtain the optimal

number of hidden states in corresponding models.

VI. DISCUSSIONS

A. Presence of Multiple People

The current system is explicitly designed and tested for scenarios involving a single person. It should be noted that this setup aligns with the intended usage scenario of the Wi-Watch, where each Officer of the Watch (OOW) is assigned specific tasks within their designated responsibility area on the bridge. Generally, similar to most current WiFi-based sensing methods, our system currently handles only single-person situations because of the complexities of discerning signals reflected from multiple individuals within the 20 MHz bandwidth of WiFi signals. Nevertheless, with the progression of wireless communication technology, upcoming protocols offering higher frequencies and larger bandwidths, such as IEEE 802.11ay [49] [50], are anticipated to address this issue. We have already conducted preliminary experiments on multi-person activity recognition using WiFi signals with a bandwidth of 160 MHz, compliant with the IEEE 802.11ax protocol. These experiments have yielded promising results and show the potential to overcome the limitations of scenarios involving a single person.

B. Device Deployment

Displacement of devices acts as a main factor limiting the recognition accuracy of the Wi-Watch. On one hand, the human reflection signal is much weaker than the LoS signal due to the longer propagation distance and additional reflection loss. Therefore, in order to increase the reflection area of humans in the experiment, we placed the antenna 1m above the ground. On the other hand, Wi-Watch utilizes extracted AoA-Doppler information as input to the activity recognition model, and any changes in the relative position between the transceiver and the person can result in variations in AoA-Doppler pairs.

C. Environment Changes

Environmental changes can have a substantial impact on the accuracy of our system. After the completion of ship construction, the placement of equipment inside the bridge remains relatively stable, resulting in minimal alterations to the bridge environment. However, one potential source of environmental change is the significant swaying of the ship

caused by wind and waves, which can lead to drastic fluctuations in the indoor channel, and even force individuals to move. The adverse impact of such conditions on the accuracy of our system is foreseeable. It should be noted that our current experiments have been conducted in inland waters characterized by relatively calm conditions. In the future, we intend to conduct experiments on ocean-going vessels to investigate the suitability of WiFi sensing technology in maritime environments that pose greater challenges.

D. Computational and Implementation Complexity

First, the Wi-Watch system is more complex compared to most existing WiFi-based activity recognition systems. This is because the Wi-Watch system needs to counteract the adverse effects caused by the low signal-to-noise ratio environment on recognition. Therefore, a series of data processing steps have been performed, resulting in our system being more complex and computationally expensive. Second, our system requires fewer parameters to be trained compared to vision-based behavior recognition methods. Vision-based methods use frameworks such as CNN-LSTM [51], GCN [52], and Transformer [53] to learn spatial and temporal patterns, which require millions of parameters. Our in-place activities recognition task uses a basic attention-based LSTM model, which requires 169,606 trainable parameters. Therefore, our model has a lower computational and implementation complexity compared to state-of-the-art vision-based behavior recognition models. With regard to the system's response time, the time consumption of our model has mainly been attributed to CSI preprocessing, the 2D-MUSIC algorithm, foreground extraction, and in-place activity recognition. According to our tests, when the transmission rate is 200 packets/s, the system's response time is 0.83 seconds. Given the need for vigilance evaluation during ship driving, this response speed is deemed acceptable.

VII. Conclusion

We present Wi-Watch, an innovative system that leverages commercial off-the-shelf WiFi devices to enable activity recognition and vigilance detection for ship-bridge watchkeeping officers. To realize this system, we first propose a novel signal path parameter estimation approach based on the CSI path model and 2D-MUSIC. To mitigate the noise introduced by ship vibrations in the WiFi signal, we introduce a sophisticated foreground extraction method that effectively separates signals reflected by the human body from those reflected by environmental objects. Subsequently, we introduce CSI velocity and activity models to recognize watchkeeping states and activities accurately. Lastly, we employ an advanced vigilance estimation model based on the duration of low-vigilance states to evaluate the levels of vigilance exhibited by watchkeeping officers during their shifts. The experimental results validate the robustness of our system, showcasing its ability to effectively mitigate the impact of low signal-to-noise ratio ship environments on WiFi signals. Our system demonstrates exceptional performance in accurately identifying watchkeeping states and activities and providing reliable vigilance level assessments.

Anticipating future research scopes, we delineate two encouraging trajectories. In our research plan, the main objective is the development of Wi-Watch, which is based on the advanced IEEE 802.11ay protocol. The superior multipath resolution and Doppler frequency resolution capabilities that this protocol offers are set to considerably enhance Wi-Watch's capacity for discerning the actions of numerous individuals and detecting subtle movements. Additionally, though our present experiments are confined to inland vessels, we anticipate widening the scope to maritime vessels as a strategy to understand potential challenges unique to that domain.

References

[1] M. Accident, I. Branch, F. Floor, C. House, C. Place, and U. Kingdom, "Bridge Watchkeeping Safety Study," no. July, 2004.

[2] S. Fan, J. Zhang, E. Blanco-Davis, Z. Yang, J. Wang, and X. Yan, "Effects of seafarers' emotion on human performance using bridge simulation," *Ocean Engineering*, vol. 170, no. October, pp. 111–119, 2018.

[3] T. A. Johansen, T. Perez, and A. Cristofaro, "Ship collision avoidance and colregs compliance using simulation-based control behavior selection with predictive hazard assessment," *IEEE transactions on intelligent transportation systems*, vol. 17, no. 12, pp. 3407–3422, 2016.

[4] S. Fan and Z. Yang, "Towards objective human performance measurement for maritime safety: A new psychophysiological data-driven machine learning method," *Reliability Engineering System Safety*, vol. 233, p. 109103, 2023.

[5] Y. Wang, X. Shi, and D. Xu, "Relationship between overconfidence and risky behavior among ship crew," *Transportation research record*, vol. 2674, no. 9, pp. 500–510, 2020.

[6] L. P. Perera, "Autonomous ship navigation under deep learning and the challenges in colregs," in *ASME 2018 37th International Conference on Ocean, Offshore and Arctic Engineering*. American Society of Mechanical Engineers Digital Collection, 2018.

[7] H. Huang, H. Chen, and S. Lin, "Magtrack: Enabling safe driving monitoring with wearable magnetics," in *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*, 2019, pp. 326–339.

[8] V. P. Balam and S. Chinara, "Statistical channel selection method for detecting drowsiness through single-channel eeg-based bci system," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–9, 2021.

[9] A. El Khatib, C. Ou, and F. Karray, "Driver Inattention Detection in the Context of Next-Generation Autonomous Vehicles Design: A Survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 11, pp. 1–14, 2019.

[10] X. Xu, J. Yu, Y. Chen, Y. Zhu, L. Kong, and M. Li, "Breathlistener: Fine-grained breathing monitoring in driving environments utilizing acoustic signals," in *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*, 2019, pp. 54–66.

[11] E. M. Nowara, T. K. Marks, H. Mansour, and A. Veeraraghavan, "Near-infrared imaging photoplethysmography during driving," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 4, pp. 3589–3600, 2022.

[12] Y. Xing, C. Lv, H. Wang, D. Cao, E. Velenis, and F. Y. Wang, "Driver activity recognition for intelligent vehicles: A deep learning approach," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 6, pp. 5379–5390, 2019.

[13] A. Koesdwiady, R. Soua, F. Karray, and M. S. Kamel, "Recent trends in driver safety monitoring systems: State of the art and challenges," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 6, pp. 4550–4563, 2017.

[14] Z. Wang, J. Hu, R. Lv, J. Wei, Q. Wang, D. Yang, and H. Qi, "Personalized privacy-preserving task allocation for mobile crowdsensing," *IEEE Transactions on Mobile Computing*, vol. 18, no. 6, pp. 1330–1341, 2019.

[15] Z. Wang, X. Pang, J. Hu, W. Liu, Q. Wang, Y. Li, and H. Chen, "When mobile crowdsensing meets privacy," *IEEE Communications Magazine*, vol. 57, no. 9, pp. 72–78, 2019.

[16] W. B. III. ICAO is Not Calling for Cockpit Video Cameras in New Airplanes - Avionics International (aviationtoday.com). (2017, March 28). [Online]. Available: https://www.aviationtoday.com/2017/03/28/icao-not-calling-cockpit-video-cameras-new-airplanes/

[17] A. Sternstein. Pilots resist FAA rule on cabin surveillance – FCW. (2005, December 28). [Online]. Available: https://fcw.com/workforce/2005/12/pilots-resist-faa-rule-on-cabin-surveillance/210171/

[18] K. Liu, Z. Tian, Z. Li, and M. Zhou, "Rfloc: A reflector-assisted indoor localization system using a single-antenna ap," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–16, 2021.

[19] M. Chen, K. Liu, J. Ma, Y. Gu, Z. Dong, and C. Liu, "Swim: Speed-aware wifi-based passive indoor localization for mobile ship environment," *IEEE Transactions on Mobile Computing*, 2019.

[20] Y. Jin, Z. Tian, H. Wang, and M. Zhou, "Wisen: Zero-knowledge passive human tracking using a single wifi link," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–15, 2022.

[21] S. Ji, Y. Xie, and M. Li, "Sifall: Practical online fall detection with rf sensing." New York, NY, USA: Association for Computing Machinery, 2023. [Online]. Available: https://doi.org/10.1145/3560905.3568517

[22] R. Gao, W. Li, Y. Xie, E. Yi, L. Wang, D. Wu, and D. Zhang, "Towards robust gesture recognition by characterizing the sensing quality of wifi signals," vol. 6, no. 1, 2022.

[23] M. Chen, J. Ma, X. Zeng, K. Liu, M. Chen, K. Zheng, and K. Wang, "Md-alarm: A novel manpower detection method for ship bridge watchkeeping using wi-fi signals," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–13, 2022.

[24] S. Zhang, W. Wang, and T. Jiang, "Wi-fi-inertial indoor pose estimation for microaerial vehicles," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 5, pp. 4331–4340, 2021.

[25] F. Xiao, S. Zhang, S. Tang, S. Shen, H. Dong, and Y. Zhong, "Wision: Bolstering mav 3d indoor state estimation by embracing multipath of wifi," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 1, pp. 253–266, 2023.

[26] K. Liu, D. Pei, S. Zhang, X. Zeng, K. Zheng, C. Li, and M. Chen, "Wicrew: Gait-based crew identification for cruise ships using commodity wifi," *IEEE Internet of Things Journal*, vol. 10, no. 8, pp. 6960–6972, 2023.

[27] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti, "SpotFi : Decimeter Level Localization Using WiFi," *Sigcomm 2015*, pp. 269–282, 2015. [Online]. Available: Stanford University

[28] X. Zeng, K. Liu, J. Ma, M. Chen, and M. Yu, "Reliability and delay trade-off analysis of unslotted ieee 802.15.4 sensor network for shipboard environment," *IEEE Sensors Journal*, vol. 21, no. 2, pp. 2400–2411, 2021.

[29] M. Chen, K. Liu, J. Ma, X. Zeng, Z. Dong, G. Tong, and C. Liu, "Moloc: Unsupervised fingerprint roaming for device-free indoor localization in a mobile ship environment," *IEEE Internet of Things Journal*, vol. 7, no. 12, pp. 11 851–11 862, 2020.

[30] H. Zhang, X. Huang, D. Pei, K. Liu, M. Chen, and X. Zeng, "Wifi based vigilant activity recognition for ship-bridge watchkeeping officers," in *International Conference on Marine Equipment & Technology and Sustainable Development*. Springer, 2023, pp. 970–977.

[31] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Predictable 802.11 packet delivery from wireless channel measurements," *ACM SIGCOMM Computer Communication Review*, vol. 40, p. 159, 2010.

[32] Y. Xie, Z. Li, and M. Li, "Precise power delay profiling with commodity wifi," in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 53–64.

[33] R. Roy and T. Kailath, "Esprit-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on acoustics, speech, and signal processing*, vol. 37, no. 7, pp. 984–995, 1989.

[34] K. Zheng, X. Zhang, X. Li, P. Li, X. Chang, J. Sang, M. Ge, and H. Schuh, "Mitigation of unmodeled error to improve the accuracy of multi-gnss ppp for crustal deformation monitoring," *Remote Sensing*, vol. 11, no. 19,

p. 2232, 2019.

[35] D. Zhang, Q. Liu, L. Chen, W. Xu, and K. Wang, "Multi-layer based multi-path routing algorithm for maximizing spectrum availability," *Wireless Networks*, vol. 24, no. 3, pp. 897–909, 2018.

[36] P. Stoica and A. Nehorai, "Music, maximum likelihood, and cramer-rao bound," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 5, pp. 720–741, 1989.

[37] K. Zheng, X. Zhang, P. Li, X. Li, M. Ge, F. Guo, J. Sang, and H. Schuh, "Multipath extraction and mitigation for high-rate multi-gnss precise point positioning," *Journal of Geodesy*, vol. 93, no. 10, pp. 2037–2051, 2019.

[38] Y. Zheng, Y. Zhang, K. Qian, G. Zhang, Y. Liu, C. Wu, and Z. Yang, "Widar3.0:Zero-effort cross-domain gesture recognition with Wi-Fi," *MobiSys 2019 - Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*, pp. 313–325, 2019.

[39] Z. Yang, Y. Zhang, G. Chi, and G. Zhang, "Hands-on wireless sensing with wi-fi: A tutorial," *arXiv preprint arXiv:2206.09532*, 2022.

[40] K. Jordan, J. H. Challis, and K. M. Newell, "Walking speed influences on gait cycle variability," *Gait Posture*, vol. 26, no. 1, pp. 128–134, 2007.

[41] R. W. Bohannon and A. Williams Andrews, "Normal walking speed: a descriptive meta-analysis," *Physiotherapy*, vol. 97, no. 3, pp. 182–189, 2011.

[42] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-home gesture recognition using wireless signals," in *Proceedings of the 19th Annual International Conference on Mobile Computing Networking*, ser. MobiCom '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 27–38.

[43] K. Qian, C. Wu, Y. Zhang, G. Zhang, Z. Yang, and Y. Liu, "Widar2.0: Passive human tracking with a single wi-fi link," in *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 350–361.

[44] D. Bonefačić and J. Bartolić, "Design considerations of an active integrated antenna with negative resistance transistor oscillator," *Radioengineering*, vol. 14, no. 4, pp. 33–39, 2005.

[45] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and Modeling of WiFi Signal Based Human Activity Recognition," *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking - MobiCom '15*, pp. 65–76, 2015.

[46] T. Hang, Y. Zheng, K. Qian, C. Wu, Z. Yang, X. Zhou, Y. Liu, and G. Chen, "Wish: Wifi-based real-time human detection," *Tsinghua Science and Technology*, vol. 24, no. 5, pp. 615–629, 2019.

[47] Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang, and H. Liu, "E-eyes: Device-free Location-oriented Activity Identification Using Fine-grained WiFi Signatures," *Proceedings of the 20th annual international conference on Mobile computing and networking - MobiCom '14*, pp. 617–628, 2014.

[48] Y. Gu, J. Zhan, Y. Ji, J. Li, F. Ren, and S. Gao, "Mosense: An rf-based motion detection system via off-the-shelf wifi devices," *IEEE Internet of Things Journal*, vol. 4, no. 6, pp. 2326–2341, 2017.

[49] L. Polak, J. Milos, R. Zedka, J. Blumenstein, and C. Mecklenbräuker, "Ber and throughput performances of ieee 802.11 ay sc-phy over measured 60 ghz indoor channels," *Telecommunication Systems*, vol. 80, no. 4, pp. 573–587, 2022.

[50] Y. Ghasempour, C. R. C. M. da Silva, C. Cordeiro, and E. W. Knightly, "Ieee 802.11ay: Next-generation 60 ghz communication for 100 gb/s wi-fi," *IEEE Communications Magazine*, vol. 55, no. 12, pp. 186–192, 2017.

[51] K. Muhammad, Mustaqeem, A. Ullah, A. S. Imran, M. Sajjad, M. S. Kiran, G. Sannino, and V. H. C. de Albuquerque, "Human action recognition using attention based lstm network with dilated cnn features," *Future Generation Computer Systems*, vol. 125, pp. 820–830, 2021.

[52] P. Zhang, C. Lan, W. Zeng, J. Xing, J. Xue, and N. Zheng, "Semantics-guided neural networks for efficient skeleton-based human action recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[53] Y. Zhang, B. Wu, W. Li, L. Duan, and C. Gan, "Stst: Spatial-temporal specialized transformer for skeleton-based action recognition," in *Proceedings of the 29th ACM International Conference on Multimedia*, ser. MM '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 3229–3237.

Danei Gong received the B.Eng. degree in communication engineering from Wuhan University of Technology (WUT), Wuhan, China, in 2020. He received M.S. degree in transportation from WUT in 2023. He is currently a research assistant with the City University of Hong Kong. His research interests are in wireless sensing.

Kezhong Liu received the B.S. and M.S. degrees in marine navigation from the Wuhan University of Technology(WUT), Wuhan, China, in 1998 and 2001, respectively. He received the Ph.D. degree in communication and information engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2006. He is currently a professor with School of Navigation, WUT. His active research interests include indoor localization technology and data mining for ship navigation.

Dashuai Pei received his B.S. degree in Automation from Northwest Minzu University in Lanzhou, China in 2018, and his M.S. degree in Transportation Engineering from Wuhan University of Technology (WUT), China, in 2022, and he is currently working towards his PhD. His research interests include wireless sensing, driver behavior recognition, and ship navigation safety.

Haoran Zhang received the B.S. degree in Electronic Science and Technology from the University of Electronic Science and Technology of China, Chengdu, China, in 2019. He is currently working toward an M.S. degree at the Wuhan University of Technology, Wuhan, China. His research interests are in wireless sensing.

Shengkai Zhang received his Ph.D. degree from the EIC Department, Huazhong University of Science and Technology (HUST) in 2021. He received his M.Sc and M.Phil degrees from HUST and Hong Kong University of Science and Technology in 2012 and 2014, respectively. He is currently an associate professor at Wuhan University of Technology (WUT). His recent research interests include state estimation, wireless sensing, mobile computing, multi-sensor fusion, robot control and planning.

Mozi Chen received a B.S. degree in electric engineering from the Hubei University of Technology, China, in 2013, an M.S. degree in navigation engineering from the Wuhan University of Technology (WUT), China, in 2016 and, and the Ph.D. degree from the WUT, China, in 2020. He is currently an associate researcher at WUT. His research work has been focused on wireless sensing techniques and machine learning algorithms for human localization, emergency navigation, and activity recognition in mobile environments, i.e., cruise ships.