IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 8 March 2024, accepted 25 April 2024, date of publication 29 April 2024, date of current version 17 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3395173

RESEARCH ARTICLE

Wi-MIR: A CSI Dataset for Wi-Fi Based Multi-Person Interaction Recognition

MD. SHAFIQUL ISLAM[1], M. HUMAYUN KABIR[2], (Member, IEEE), MD. ALI HASAN[3], AND WONJAE SHIN[1], (Senior Member, IEEE)

[1]School of Electrical Engineering, Korea University, Seoul 02841, Republic of Korea
[2]Department of Electrical and Electronic Engineering, Islamic University, Kushtia 7003, Bangladesh
[3]Department of AI Convergence Network, Ajou University, Suwon-si 16499, Republic of Korea

Corresponding author: Wonjae Shin (wjshin@korea.ac.kr)

This work was supported in part by the National Research Foundation of Korea (NRF) under Grant 2022R1A2C4002065, and in part by the Institute of Information & Communications Technology Planning & Evaluation (IITP) under Grant 2021-0-00467 and Grant 2022-0-00704.

ABSTRACT Wi-Fi based human activity recognition is emerging and preferable over other approaches due to its numerous advantages, including privacy considerations, ubiquitousness, and easy deployment. While existing literature predominantly focuses on identifying the activities of a single-user, recognizing multi-person interactions (MPIs) is increasingly significant due to their profound social implications. However, research in this area has not progressed due to the limitation of publicly available Wi-Fi datasets and the complexities of MPI recognition. Motivated by this, we develop and publicly release a Wi-Fi channel state information (CSI) based MPI recognition dataset, coined *Wi-MIR*, that uses three transmit and three receive antennas to capture 270 ($3 \times 3 \times 30$) subcarriers with a sampling rate of 950 Hz. This dataset consists of 3,740 trials encompassing seventeen distinct MPIs and is collected by eleven human pairs in an indoor environment. We put forth a lightweight deep learning model with attention mechanisms for MPI recognition from CSI, named CSI-IRNet, that adeptly concentrates on pertinent features, filtering out irrelevant elements, and mitigating the impact of signal complexity within the CSI for recognizing MPIs accurately. In addition, we compare the developed Wi-MIR and the existing public dataset by evaluating the performance of MPI recognition on both datasets to highlight the strengths and advancements provided by Wi-MIR. The evaluation results show that Wi-MIR dataset demonstrates a superior recognition performance by utilizing more subcarriers with a higher sampling rate as well as covering more diverse kinds of MPIs (Bowing, Conversation, Exchanging objects, Helping standup, Helping walk, and Touching another person).

INDEX TERMS Channel state information, deep learning, machine learning, multi-person interaction, public dataset, received signal strength indicator, Wi-Fi.

I. INTRODUCTION

Human Activity Recognition (HAR) refers to the procedure of determining specific human behavior by analyzing a sequence of observations derived from human actions with contextual surroundings. Its applications extend to the Internet of Things (IoT), fostering the intuitive interactions between humans and machines [1], [2]. In the realm of IoT, HAR utilizes interconnected devices and sensors to monitor and track human behavior for purposes such as enhancing health services [3], ensuring security [4], and providing entertainment solutions [5], [6] through the collection, transmission, and analysis of relevant data. By leveraging IoT technology, HAR contributes to a better understanding of human activities, enhancing the capabilities of technology to adapt and respond effectively. Traditional HAR methodologies employ various sensing technologies, including computer vision [7], wearable sensors [8], and radars [9]. However, vision-based HAR that uses cameras to capture human movement data is susceptible to factors, such as varying light conditions, environmental settings, and harms user privacy. Wearable sensor-based HAR relies on users

The associate editor coordinating the review of this manuscript and approving it for publication was Byung-Seo Kim.

carrying devices, like smartphones or smartwatches, which poses inconveniences, especially for elderly or disabled individuals. The need for users to have the required tools with them can make activity recognition (AR) impractical in some real-life scenarios. Radar-based HAR, while effective, often faces limitations due to the use of specialized equipment, like universal software radio peripherals (USRP) and restricted coverage area. All these limitations collectively impact the overall performance and applicability of conventional HAR methods in various scenarios.

In contrast to conventional methods, the recent attention [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20] is garnered by Wi-Fi based HAR in indoor environments, highlighting its compatibility with the IoT. Wi-Fi based HAR stands out for its simplicity, contactless nature, non-line-of-sight (NLOS) communication, and privacy-preserving features. The contactless nature of signals is advantageous, eliminating the need for individuals to wear or carry specific devices. The inherent advantages of Wi-Fi based HAR, coupled with the principles of IoT, create a synergistic relationship that makes a promising and practical choice for AR in different contexts.

Multi-person interaction (MPI) recognition is essential to enhance collaboration, improving communication in group sitting, and coordination across various domains. It plays an important role in improving teamwork, emergency response, education, healthcare, entertainment, and the development of smart environments, leading to more effective and personalized solutions in diverse real-life scenarios [21], [22]. However, in the existing literature, authors [18], [19], [20], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33] mostly focus on developing Wi-Fi based single-user AR datasets. The research [10], [18], [20], [34], [35], [36] efforts have concentrated on developing algorithms and models that can accurately identify and classify individual activities. These AR systems have inherent limitations in real-world applications, where human interactions inherently involve more than one participant. From the literature, there is a predominant emphasis on single-user AR rather than the investigation of MPI recognition. The limitation of MPI recognition can be attributed to two primary reasons: i) a scarcity of available Wi-Fi based MPI datasets hinders the development of algorithms or models for recognizing MPIs, and ii) the complexity of analyzing MPIs between multiple individuals that involves understanding the dynamics, dependencies, and diverse scenarios in MPIs.

With the rapid advancement of deep learning (DL) across diverse domains, researchers increasingly employ machine learning (ML) and DL algorithms in Wi-Fi based HAR. Despite the commendable efforts of a few previous works [15], [16], [37], [38], [39], [40] in utilizing DL for MPI recognition with satisfactory accuracy, the research contends that a more extensive dataset can significantly enhance performance across diverse MPIs in a variety of indoor environments. There is no Wi-Fi based public dataset for MPI recognition except the existing Wi-HHI [40] dataset concerning human-to-human interaction (HHI) recognition shared by Jordan University of Science and Technology in 2020. Moreover, the inconsistency in MPI data poses a challenge for comparing related works, prompting us to construct a public Wi-Fi based MPIs dataset. Motivated by this, we introduce Wi-MIR, a dataset for Wi-Fi CSI-based MPI recognition. This dataset utilizes three transmit and three receive antennas to capture 270 ($3 \times 3 \times 30$) subcarriers with a higher sampling rate of 950 Hz. Eleven human pairs are involved in collecting 3,740 trials of seventeen distinct MPIs, including five new types of MPIs (Bowing, Conversation, Exchanging objects, Helping standup, Helping walk, and Touching another person) in an indoor environment. The developed Wi-MIR dataset addresses the current scarcity of resources and is a crucial contribution for researchers seeking to explore MPIs in diverse contexts. The advantages of constructing a Wi-Fi CSI-based public Wi-MIR dataset are threefold. Firstly, it mitigates the time and labor cost of gathering such extensive data. Secondly, it facilitates the sharing of substantial MPI data, fostering collaboration and knowledge exchange within the research community. Finally, the availability of the developed Wi-MIR dataset will accelerate the development of wireless sensing technologies and advance practical applications in MPI recognition.

To validate the worthiness of the developed Wi-MIR dataset for the effective classification of MPIs, we propose a DL-based CSI-IRNet model comprising two phases: feature extraction and classification. The feature extraction phase utilizes the depthwise separable convolution (DS-Conv) and the frame subcarrier antenna attention mechanism (FSAAM) block to extract robust features from Wi-Fi channel state information (CSI). Subsequently, these fine-grained features are used in the classification phase to recognize MPIs effectively. Moreover, the developed Wi-MIR dataset is also evaluated with the existing Wi-Fi CSI-based DL models and four ML classifiers. The contributions of this paper include:

- We develop and publicly release a Wi-Fi CSI-based MPIs dataset, named *Wi-MIR* [41], for future research of MPI recognition and the development of various human-computer interaction (HCI) applications. The developed Wi-MIR dataset consists of seventeen well-known MPIs. By leveraging the 5 GHz band with three transmit antennas ($N_{\mathsf{Tx}} = 3$), three receive antennas ($N_{\mathsf{Rx}} = 3$), along with 30 subcarriers ($N_{\mathsf{Sc}} = 30$) and a higher sampling rate of 950 Hz, the developed Wi-MIR dataset recorded 270 ($3 \times 3 \times 30$) subcarriers that significantly contribute to ensuring more accurate and robust recognition capabilities for MPIs.
- We put forth CSI-IRNet, a lightweight DL model that uses DS-Conv and FSAAM block to improve the MPI recognition using essential features from the CSI signals of three transmit and three receive antennas.
- To evaluate the developed Wi-MIR dataset, we employ both ML and DL methodologies. The evaluation incorporates four ML classifiers, two existing DL models, and the proposed CSI-IRNet model.

TABLE 1. Public datasets for Wi-Fi based AR and HHIs.

Dataset	No. of activities	Recognition type	Person involved in each activity	Data processing	Feature extraction methods	models
Wei *et al.* [23]	8	human activity	1	performed	Not provided	SRC
Wang *et al.* [24]	9	human activity	1	performed	STFT, DWT	CARM
Zheng *et al.* [25]	6	human activity	1	performed	Temporal and frequency correlation	Periodically analysis (autocorrelation)
Yousefi *et al.* [20]	7	human activity	1	performed	STFT, DWT	SVM, RF, LSTM
Jiang *et al.* [26]	6	human activity	1	performed	Not provided	RF, CNN, AN
Guo *et al.* [19]	13	human activity	1	performed	Not provided	SVM, RF, CNN, LSTM
Schäfer *et al.* [27]	8	human activity	1	performed	DWT	SVM, LSTM
Zhang *et al.* [29]	9	human activity	1	performed	Not provided	DSM
Chen *et al.* [18]	10	human activity	1	Not provided	Not provided	RF, LSTM, ABLSTM
Wang *et al.* [30]	6	human activity	1	performed	Not provided	SVM, CNN
Galdino *et al.* [33]	17	human activity	1	performed	DTW	Naive Bayes, SVM, RF
Alsaify *et al.* [31]	8	human activity	1	Not provided		
Meneghello *et al.* [32]	7	human activity	1	Not provided		
Alazrai *et al.* [40]	12	HHIs	2	Not provided		

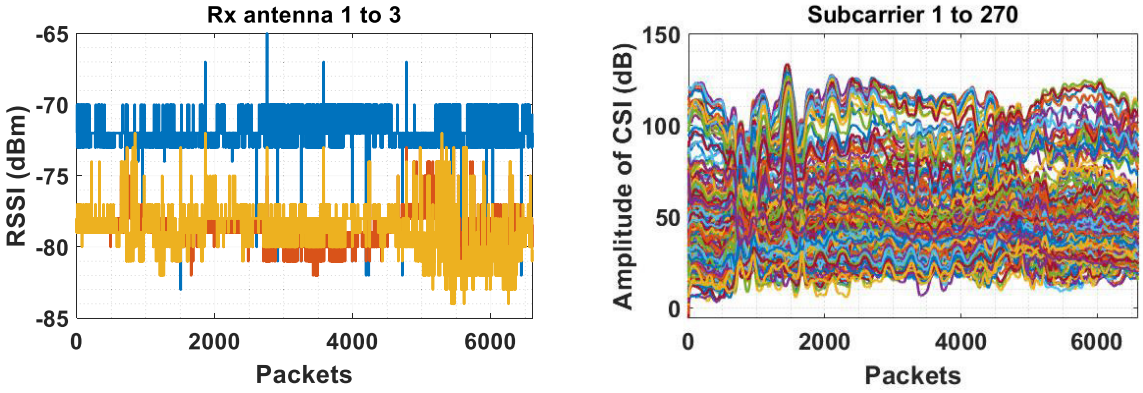
- To assess the efficacy of the developed Wi-MIR dataset compared to the existing Wi-HHI [40] dataset, we utilize the ML classifiers, the existing DL, and the proposed CSI-IRNet models. The performance results of the developed Wi-MIR dataset reveal its reliability, and the online availability ensures its usability for other researchers.

II. RELATED WORKS

There are numerous HAR datasets available in the literature, encompassing varied data collection modalities, such as cameras, wearable sensors, and Wi-Fi technology. With the increasing availability and numerous advantages of Wi-Fi communication, there is a recent upswing in the popularity of Wi-Fi based HAR datasets and methodologies. This section thoroughly reviews existing datasets specifically oriented toward Wi-Fi, delineating their significance in the ongoing exploration of AR and HHIs within the research domain.

Wei et al. [23] develop a dataset for single-user AR based on Wi-Fi and propose a sparse representation classification (SRC) method. Their proposed SRC approach is designed to enhance the robustness of location-oriented AR by leveraging CSI. In [24], authors present a Wi-Fi CSI-based single-user AR dataset and introduce a CSI-based HAR and monitoring (CARM) system. It utilizes two models to link the correlation between CSI value dynamics and specific human activity. The CARM demonstrates robust classification performance for nine activities. Smokey, a passive smoking detection system proposed in [25], utilizes the variation of Wi-Fi CSI signals. It employs a foreground detection-based motion acquisition method (temporal and frequency correlation) to extract meaningful information from multiple subcarriers. By leveraging common features based on autocorrelation, Smokey accurately recognizes smoking-related motions. Yousefi et al. [20] introduce a Wi-Fi based single-user AR dataset consisting of seven activities and employ short-time Fourier transform (STFT) and discrete wavelet transform (DWT) for features extraction. The activities are classified using support vector machine (SVM), random forest (RF), and long-term short memory (LSTM). A device-free single-user AR system is proposed in [26] using an adversarial network (AN). They collect six human activities of CSI data from six different environments. The proposed system consists of a convolutional neural network (CNN) based feature extractor, an activity recognizer, and a domain discriminator to recognize environment-independent human activities.

A public Wi-Fi based single-user AR dataset, named Wiar [19], is developed by featuring sixteen activities in three indoor environments. Various ML and DL approaches are employed to classify activities in indoor environments. Schäfer et al. [27] use the open-source tool Nexmon with Raspberry Pi and Asus RT-AC86U routers to collect a single-user AR dataset. They collect eight kinds of activity data and employ LSTM and SVM to classify human activities. In another approach [29], a single-user AR dataset is developed consisting of nine kinds of human activities. The diffraction-based sensing model (DSM) is introduced to assess signal changes in response to human movements. This model establishes a quantitative link between patterns of signal variation and corresponding motions to recognize human activities. Chen et al. [18] collect a single-user AR dataset that comprises seven kinds of activity data from two different environments (activity and meeting room) and propose an attention-based bi-directional LSTM (ABLSTM) for classifying the human activities. A dual-task CNN-based DL framework for single-user AR using Wi-Fi CSI signals is proposed in [30]. They collect over 1,400 CSI fingerprints of six activities at sixteen indoor locations. The proposed framework achieves robust performance in recognizing activities. The eHealth CSI dataset, developed by Galdino et al. [33], contains seventeen types of single-human activities. It also includes phenotype information and heartbeat rate monitoring data from a smartwatch. The dynamic time warping (DTW) algorithm is utilized to extract features, and the activities are classified using Naive Bayes, SVM, and RF classifiers. Meneghello et al. [32] collect a dataset in seven different environments. The data collection



(a) RSSI (b) CSI

FIGURE 1. RSSI and CSI signal for the MPI of the Hugging in the developed Wi-MIR dataset.

encompasses a bedroom, a kitchen, a university laboratory, a living room, a meeting room, a university office, and a semi-anechoic chamber. Throughout data collection, the respective positions of the transmitter, receiver, and monitoring devices are changed in a predefined order. Another publicly available single-user AR dataset is proposed in [31]. They collect seven kinds of human activity data using Wi-Fi signals in both the Line-of-Sight (LOS) and NLOS environments.

Alazrai et al. [40] develop an HHIs dataset, named Wi-HHI, in an indoor environment. They use a Sagemcom 2704 router with two antennas ($N_{\mathsf{Tx}}=2$) as a transmitter and a desktop computer equipped with an Intel 5300 Network Interface Card (NIC) with three antennas ($N_{\mathsf{Rx}} = 3$) as a receiver. They use an online Linux 802.11n CSI tool [42] to capture 180 subcarriers. The Wi-HHI [40] dataset consists of 13 HHIs: 12 distinct HHIs and one steady state. They involve 44 human pairs to collect 12 different HHIs by setting the transmitter to run at 2.4 GHz and a channel bandwidth of 20 MHz. A summary of the existing Wi-Fi based datasets of AR and HHIs is given in Table 1.

The reviewed literature highlights the creation of Wi-Fi based datasets for single-user AR by employing diverse signal processing based on ML and DL approaches. There is only one Wi-Fi CSI-based public dataset concerning HHIs. Recognizing the scarcity, effectiveness, and importance of MPIs, this study focuses on the development of a Wi-Fi CSI-based MPIs dataset. The developed Wi-MIR dataset utilizes three transmit and three receive antennas with more subcarriers and a higher sampling rate of 950 Hz. This dataset deals with seventeen distinct MPIs, including five new kinds of MPIs (Bowing, Conversation, Exchanging objects, Helping standup, Helping walk, and Touching another person). In addition, we present a DL-based model (CSI-IRNet) with an attention mechanism to evaluate the developed Wi-MIR dataset.

III. WI-FI PROPERTISE

Wi-Fi signal has two types of properties: Received Signal Strength Indicator (RSSI) and CSI. Each of these properties has unique characteristics and capabilities. Therefore, analyzing both the RSSI and CSI plays an important role in effectively recognizing MPIs.

A. RECEIVE SIGNAL STRENGTH INDICATOR (RSSI)

RSSI is the sum of signal powers from multiple paths measured at each receive antenna, including the LOS path between the transmitter and receiver, along with reflections from surfaces like walls, furniture, and human entities in the wireless communication environment. Separating the signal components produced by the various paths in an indoor environment is challenging due to the multipath effect. Moreover, RSSI is considered a coarse-grained indicator of Wi-Fi signals, broadly representing the overall received signal power. This characteristic is beneficial for capturing dynamic changes, especially in human movement scenarios. Therefore, RSSI indicates simple dynamic variations in the wireless environment. Fig. 1 (a) presents the RSSI signal for the MPI of the Hugging in the developed Wi-MIR dataset.

B. CHANNEL STATE INFORMATION (CSI)

CSI stands as a pivotal element within wireless communication systems by unraveling the intricacies of the communication channel between transmitters and receivers. This dynamic information encapsulates essential parameters, like amplitude, phase, and frequency response, that offers a detailed perspective on the behavior of the wireless channel. In the orthogonal frequency division multiplexing (OFDM) framework, the wireless channel is divided into several narrowband subcarriers. The communication system can be represented as [28]:

$$\mathbf{y}_i = \mathbf{H}_i\mathbf{x}_i + \mathbf{v},\ \ i = 1, 2, 3, \ldots, N, \tag{1}$$

where $\mathbf{H}_i \in \mathbb{C}^{N_{\mathsf{Tx}} \times N_{\mathsf{Rx}}}$ embodies the CSI matrix of the i-th subcarrier; $\mathbf{v}$ and N represent the noise term and the number of OFDM subcarriers, respectively. $\mathbf{y}_i \in \mathbb{R}^{N_{\mathsf{Rx}}}$ and $\mathbf{x}_i \in \mathbb{R}^{N_{\mathsf{Tx}}}$ are the receive and transmit signals, respectively. The CSI

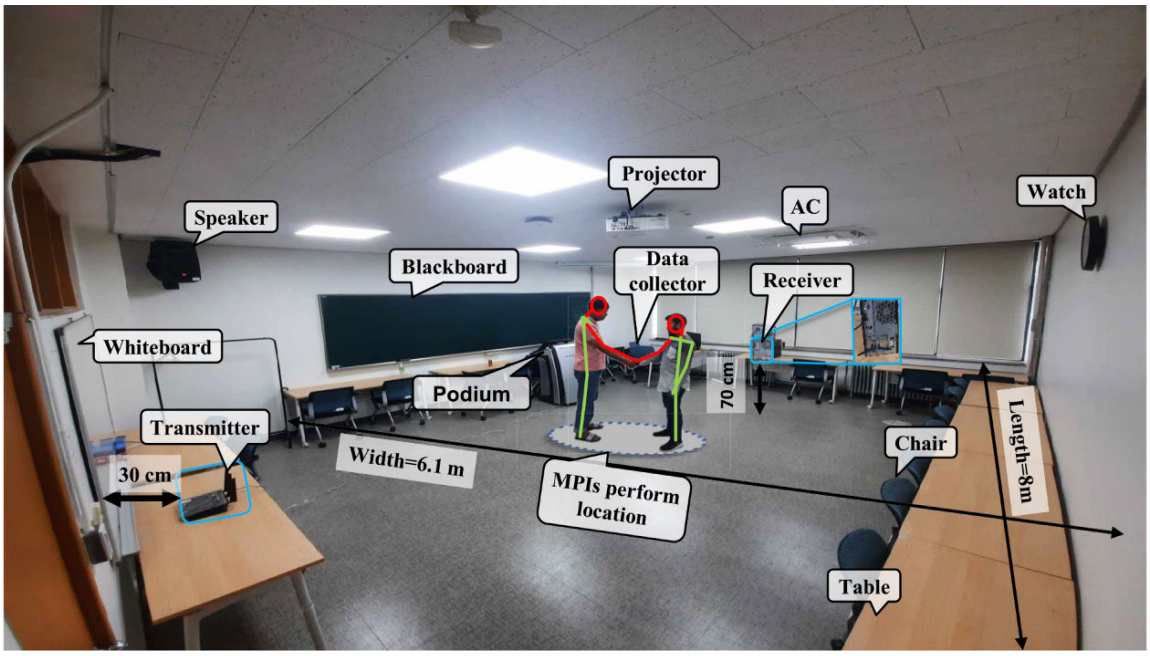


FIGURE 2. The testbed environment for MPIs data collection.

matrix entries $\mathbf{H}_i$ are detailed as:

$$\mathbf{H}_i = \begin{bmatrix} h_i^{11} & h_i^{12} & \dots & h_i^{1N_{\text{Rx}}} \\ h_i^{21} & h_i^{22} & \dots & h_i^{2N_{\text{Rx}}} \\ \vdots & \vdots & \ddots & \vdots \\ h_i^{N_{\text{Tx}}1} & h_i^{N_{\text{Tx}}2} & \dots & h_i^{N_{\text{Tx}}N_{\text{Rx}}} \end{bmatrix}. \quad (2)$$

These elements are represented as h_i^{jk} that describe the CSI of the i-th subcarrier for the link between the j-th transmit antenna and the k-th receive antenna. For a given element, h_i^{jk} is a complex value that is represented as:

$$h_i^{jk} = |h_i^{jk}| e^{j\angle h_i^{jk}}, \quad (3)$$

where $|h_i^{jk}|$ and $\angle h_i^{jk}$ represent the amplitude and phase, respectively. CSI encompasses amplitude and phase information across multiple subcarriers, providing a fine-grained view of the wireless channel. This characteristic enables better discrimination between different signal propagation paths, aiding in understanding multipath propagation, reflections, and signal interference [43]. Due to its capacity to capture rich information, CSI becomes a valuable indicator for recognizing and understanding various MPIs in a wireless environment. Fig. 1 (b) presents the CSI signal of 270 ($3 \times 3 \times 30$) subcarriers for the MPI of the Hugging in the developed Wi-MIR dataset.

IV. DEVELOPED WI-MIR DATASET

In the domains of ML and DL applications, datasets serve as the cornerstone resources. The size and quality of a dataset collectively influence the efficacy of classification or recognition models. With increasing attention to Wi-Fi based HAR technology and the limited availability of Wi-Fi based MPI datasets, it is necessary to build robust and publicly available Wi-Fi based MPI datasets. To bridge the gap, this research introduces a Wi-Fi CSI-based MPIs dataset, named Wi-MIR, by capturing seventeen diverse MPIs (Approaching, Bowing, Conversation, Departing, Exchanging objects, Handshaking, Helping standup, Helping walk, Hugging, Kicking with the left leg, Kicking with the right leg, Pointing with the left hand, Pointing with the right hand, Punching with the left hand, Punching with the right hand, Pushing, and Touching another person). Detailed insights into the developed Wi-MIR dataset, underscoring the richness and depth of the available data for analysis, are presented in the following subsections.

A. TESTBED SCENARIO

The Wi-Fi signals of seventeen MPIs are recorded in an indoor environment, specifically in a furnished room with dimensions of $8\text{m} \times 6.1\text{m}$, as depicted in Fig. 2. In Fig. 2, the red-marked area exhibits frequent movement of human limbs, while the green area tends to remain relatively stable compared to the moving area during performing MPIs. The testbed room is maintained under uniform conditions for all participants to ensure data comparability. The room is furnished with 12 tables, 24 chairs, a podium, two boards, an air conditioning (AC), a speaker, a clock, and a projector. A Wi-Fi network is established within this environment, utilizing a Wi-Fi router as an access point (AP) and a desktop computer with an Intel 5300 NIC as a receiver. The AP and Intel 5300 NIC are strategically positioned in a LOS configuration, maintaining a distance of 7.4 m between them. Throughout the study, the pair of humans engage in seventeen MPIs within the central area between the AP and the receiver by ensuring standardized conditions for reliable research outcomes and comparisons.

FIGURE 3. The AP and Intel 5300 NIC are used to transmit and receive Wi-Fi packets.

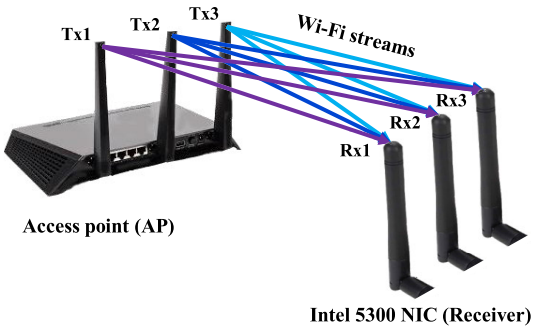


FIGURE 4. The MIMO streams established between the AP and the receiver (Intel 5300 NIC).

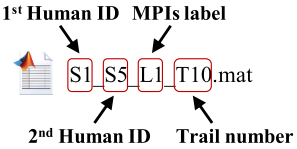


FIGURE 5. Naming format for each recorded file.

B. HARDWARE AND DEVELOPMENT TOOLS

We utilize the publicly available CSI tool [42] to record Wi-Fi signals transmitted from an AP (NETGEAR Nighthawk-R7000 router) to a receiver (a desktop computer equipped with an Intel 5300 NIC). Fig. 3 depicts the NETGEAR Nighthawk-R7000 router and the Intel 5300 NIC engaged in MPI data collection. The AP operates at a 5 GHz band, specifically on channel number 36, with a channel bandwidth of 20 MHz. The receiver is connected to the AP by a dynamically assigned IP address. The AP and receiver adhere to the IEEE 802.11n standard, incorporating multiple-input multiple-output (MIMO) technology. Both the AP and receiver are equipped with three transmit antennas ($N_{\mathsf{Tx}} = 3$) and three receive antennas ($N_{\mathsf{Rx}} = 3$) to form a MIMO system with 3×3 Wi-Fi streams. We used CSI tool [42] that captures CSI for 30 subcarriers ($N_{\mathsf{Sc}} = 30$) within each OFDM modulated MIMO stream. Therefore, this design facilitates capturing 270 ($3 \times 3 \times 30$) subcarriers in the MIMO system. Fig. 4 illustrates the MIMO streams established between the AP with three transmit antennas (Tx1, Tx2, and Tx3) and the Intel 5300 NIC with three receive antennas (Rx1, Rx2, and Rx3) that facilitate the recording of CSI during seventeen MPIs.

C. DATA FORMAT

The raw data is organized into a central folder with seventeen subfolders. Each subfolder contains data files recorded for specific MPIs. Notably, there are 3,740 recorded trials, with 20 trials for each MPI across 11 pairs (i.e., 17 MPIs $\times$ 20 trials $\times$ 11 pairs of human). Each trial is stored as a distinct MATLAB data file (.mat) with a unique name. The nomenclature for each recorded data file follows a structure of "first_second_third_fourth.mat", as depicted in Fig. 5. The first and second segments denote the pair of human IDs involved in the recorded MPIs, while the third part indicates one of the seventeen MPIs represented by an integer from 1 to 17. Specifically, Approaching (L1), Bowing (L2), Conversation (L3), Departing (L4), Exchanging objects (L5), Handshaking (L6), Helping standup (L7), Helping walk (L8), Hugging (L9), Kicking with the left leg (L10), Kicking with the right leg (L11), Pointing with the left hand (L12), Pointing with the right hand (L13), Punching with the left hand (L14), Punching with the right hand (L15), Pushing (L16), and Touching another person (L17). The remaining part indicates the trial number, ranging from 1 to 20. For instance, the data file "S1_S5_L1_T10.mat" corresponds to data recorded for a pair of human IDs, S1 and S5, during the tenth trial of the MPI of the Approaching. Each trial's data file consists of a cell array with dimensions $M \times 1$, where M represents the count of Wi-Fi packets captured during that specific trial. In addition, each element within the cell array stores a Wi-Fi packet as a structured entity, as delineated in Table 2.

D. SUMMARY OF PARTICIPANTS AND DATA ACQUISITION

In the developed Wi-MIR dataset, six males are randomly chosen to participate in seventeen predetermined MPIs. The impact of each human on the reflected signals is systematically evaluated by considering factors like height, weight, and exercise experience, as detailed in Table 3. To facilitate the collection of MPI data, we create 11 unique human pairs from the six individuals who willingly participated in this experimental study. Throughout the data collection process, a designated data collector consistently occupied a corner of the testbed room to collect MPI data.

Each pair of humans is assigned to execute 20 trials for each of the seventeen MPIs. Sample images depict these MPIs as presented in Fig. 6. The MPIs are characterized by two distinct intervals: the steady state and the MPI intervals. The pair of humans stand in a steady state interval without engaging in any specific activity. Conversely, the pair of humans execute one of the seventeen predefined MPIs during the MPI interval. We design timing diagrams outlining the sequence of the MPI to ensure precision in performing seventeen MPIs. During data collection, a start beep indicates the commencement of the steady state. A second beep indicates a pair of humans initiating one of the MPIs, while a third beep specifies the ending of each MPI trial. The timing diagram for recording is illustrated in Fig. 7.

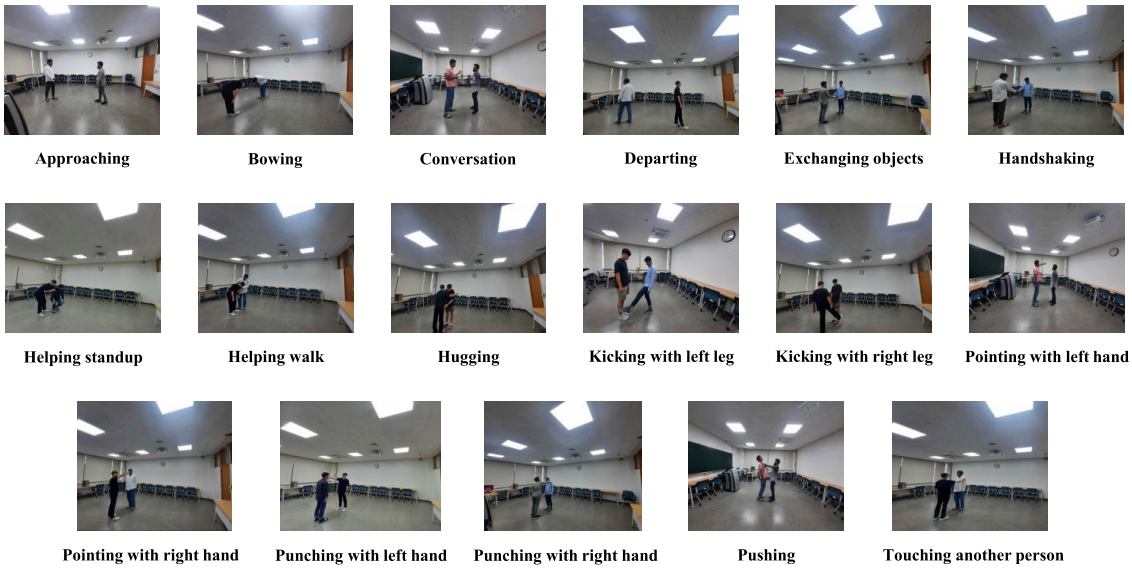


FIGURE 6. Sample images of the seventeen MPIs considered in the developed Wi-MIR dataset.

TABLE 2. The description of the fields in the structure that contains a Wi-Fi packet.

Fields	Description
N_{Tx}	Number of antennas to send the Wi-Fi packets at the transmitter.
N_{Rx}	Number of antennas to receive the Wi-Fi packets at the receiver.
noise	Represents the noise value.
RSSI_a	The RSSI value captured by the first antenna of the receiver.
RSSI_b	The RSSI value captured by the second antenna of the receiver.
RSSI_c	The RSSI value captured by the third antenna of the receiver.
CSI	Three-dimensional complex matrix with dimensions of $N_{Tx} \times N_{Rx} \times N_{Sc}$.
MPIs_label	L0, L1, L2, L3, L4, L5, L6, L7, L8, L9, L10, L11, L12, L13, L14, L15, L16, and L7.

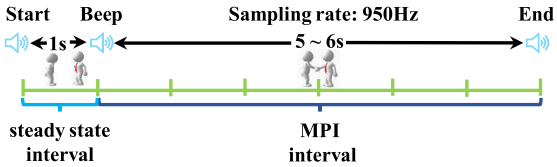


FIGURE 7. Timing diagram for MPIs data collection.

TABLE 3. Summary of the of the participant involved MPIs data collection.

Human ID	Sex	Height (cm)	Weight (kg)	Age (years)	Experience
S1	male	180	92	47	Yes
S2	male	170	62	32	Yes
S3	male	167	79	27	Yes
S4	male	178	85	25	Yes
S5	male	171	60	25	Yes
S6	male	175	72	35	Yes

E. DATASET COLLECTION LOCATION AND ACCESSIBILITY

The data for this study is collected from Ajou University, specifically from the Department of Electrical and Computer Engineering, located in Suwon, Gyeonggi-do, Republic of Korea. The geographical coordinates for the data collection site are approximately 37.2829° North latitude and 137.0435° East longitude. The developed dataset is publicly available for researchers interested in further exploration. The dataset is available on [41].

V. WI-MIR DATASET ANALYSIS

This section delves into a comprehensive analysis of the Wi-Fi CSI-based Wi-MIR dataset, consisting of seventeen distinct MPIs. The analysis is facilitated by examining various figures, each shedding light on specific aspects of the dataset. Our analysis commences with a focus on the CSI signals recorded during each of the seventeen MPIs.

Fig. 8 presents a three-dimensional mesh plot showcasing CSI signals recorded during the seventeen MPIs involving the pair of humans comprising human IDs 1 and 6 in the developed Wi-MIR dataset. The plot unfolds the temporal dynamics (revealing patterns, trends, or changes throughout the MPIs) within each MPI, visually representing two distinct intervals. The initial 1-second duration is characterized by the interval of steady state, followed by the remaining seconds dedicated to the MPI interval. The X-axis, Y-axis, and Z-axis represent time in seconds, subcarriers, and the amplitude of CSI signals, respectively. Notably, the plot captures nuances in CSI signal variations across subcarriers over time and reveals potential patterns during MPIs. This visualization offers valuable insights into the spatiotemporal dynamics of Wi-Fi signals by contributing to a comprehensive understanding of MPIs of the developed Wi-MIR dataset.

The CSI signal of three transmit and three receive antennas during the MPI of the Approaching is shown in Fig. 9. It represents the signal patterns during both the steady state

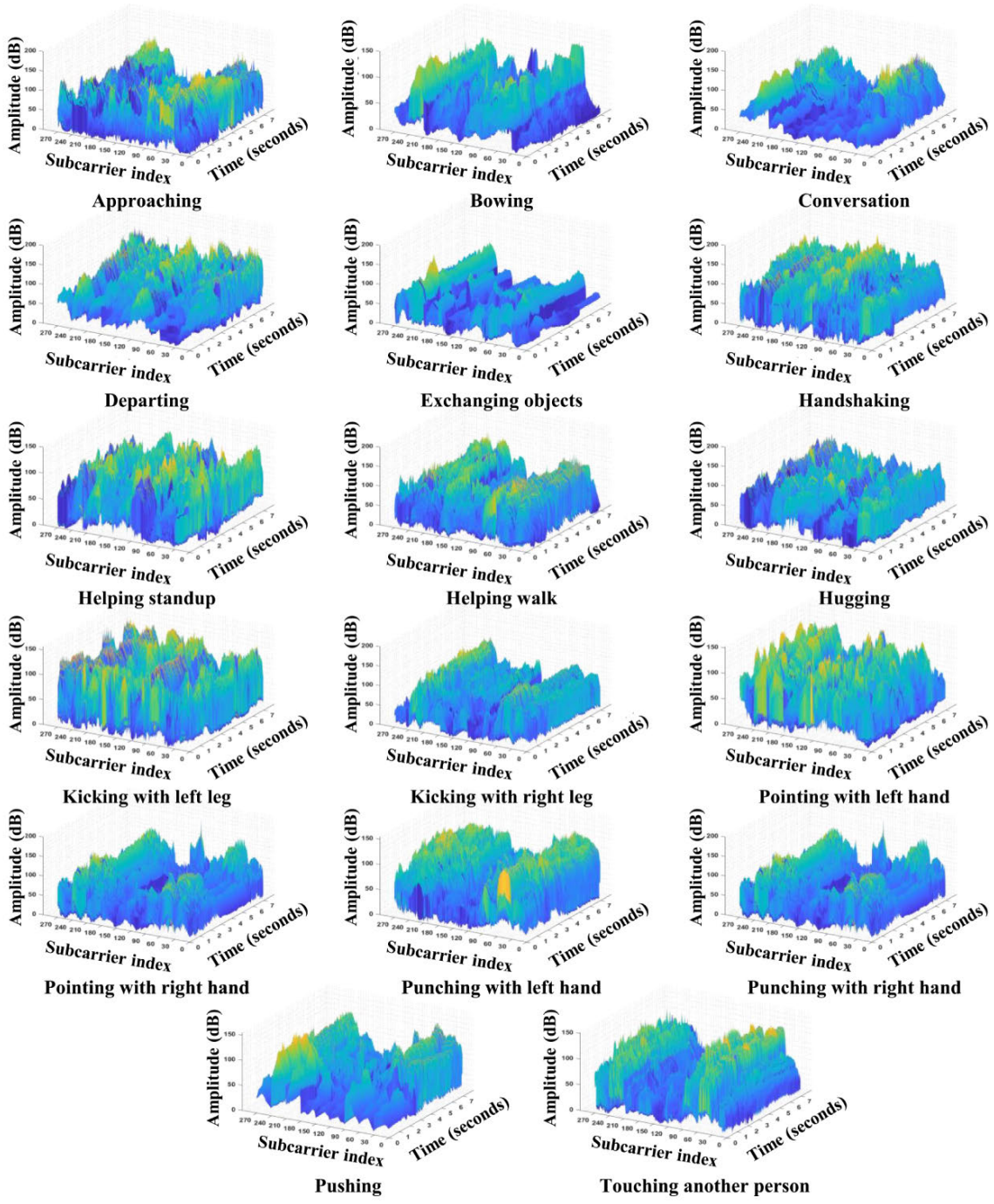


FIGURE 8. The CSI signals are captured during seventeen MPIs involving the pair of humans comprising human IDs 1 and 6 in the developed Wi-MIR dataset.

and MPI intervals. The signal patterns within the MPI interval change rapidly, whereas signals associated with the interval of the steady state exhibit a more relaxed trend. Moreover, it highlights that the signal patterns vary across three transmit antennas and three receive antennas for the same MPI. It emphasizes the need for a detailed analysis of each MPI with corresponding changes in transmit-receive antenna pairs that can be helpful in extracting fine-grained features in MPI recognition.

Fig. 10 presents the average ± standard deviation of Wi-Fi packets across all intervals, including MPIs and steady state intervals, for each of the seventeen MPIs. We can see that every MPI comprises an average of 6,500 packets for all intervals (MPIs and steady state), where 950 packets are

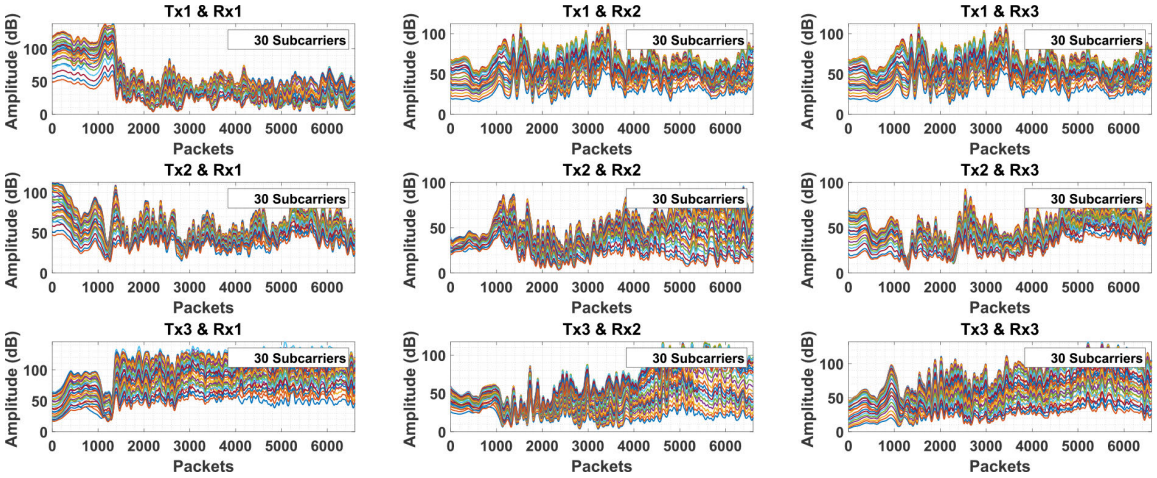


FIGURE 9. The CSI signals of 270 (3 × 3 × 30) subcarriers for the MPI of the approaching among three transmit and three receive antennas.

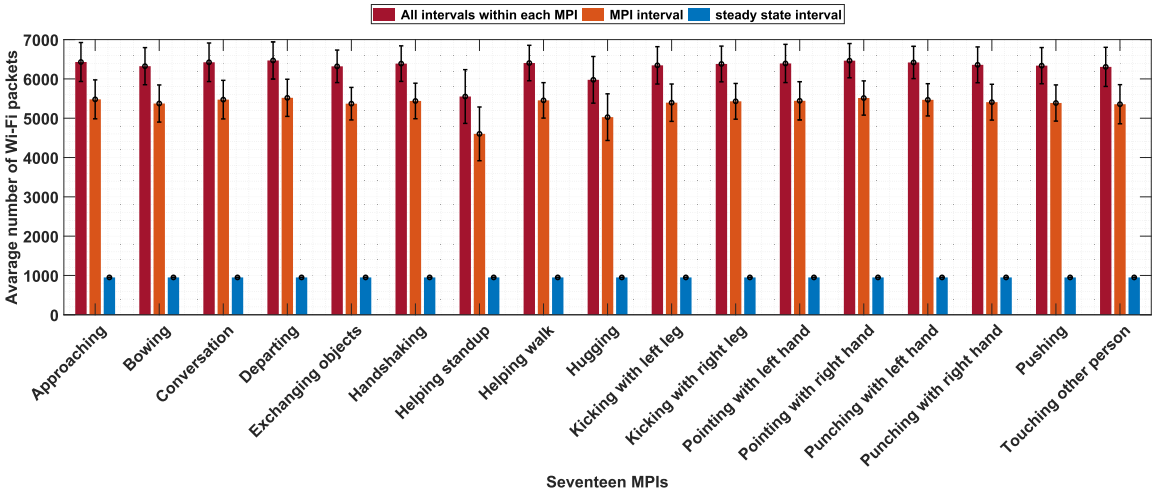


FIGURE 10. The summary of the recorded data packets of all the seventeen MPIs.

for the steady state and the remaining packets for the MPIs interval.

VI. MULTI-PERSON INTERACTION RECOGNITION METHODOLOGY

The schematic representation of the proposed MPI recognition system is illustrated in Fig. 11. This framework encapsulates the pivotal stages essential for accurate MPI recognition. Comprising four integral components, the proposed framework unfolds as follows: i) MPIs data collection, ii) Raw CSI data preprocessing, iii) Splitting the preprocessed data into ten folds for training and testing, and iv) Training, validation, and evaluation of the proposed CSI-IRNet model. In our work, we collect 17 MPIs data, and a comprehensive overview of the data collection process is provided in Section IV. Noise is induced while the Wi-Fi signal propagates, so we employ the widely used low-pass Butterworth filter to denoise the raw CSI signals. Finally, we propose a CSI-IRNet model to extract robust features and recognize MPIs accurately.

A. DATA PREPROCESSING

The raw CSI data is collected as a four-dimensional tensor encompassing packet index (time), OFDM subcarriers, and spatial variations between transmit-receive antenna pairs. Due to interventions like signal attenuation, multipath propagation, and environmental factors, Wi-Fi CSI-based signals often contain noise and outliers. It is essential to preprocess the raw CSI data by removing irrelevant information to enhance the credibility of MPI-related signal patterns. We utilize a commonly used low-pass Butterworth filter [10], [15] with a cutoff frequency of 10 Hz to denoise the raw CSI signal, and this filtering process occurs within the four-dimensional data space. The data has dimensions $I \times N_{\mathsf{Rx}} \times N_{\mathsf{Tx}} \times N_{\mathsf{Sc}}$, where $N_{\mathsf{Sc}} = 30$, $N_{\mathsf{Tx}} = 3$, and $N_{\mathsf{Rx}} = 3$ represent number of subcarriers, transmit antennas, and receive antennas, respectively; I denotes the number of packets recorded during a specific trial. After denoising, the resulting four-dimensional filtered CSI data undergoes the transformation into a two-dimensional matrix with dimensions $I \times D$, that retains temporal, spectral, and

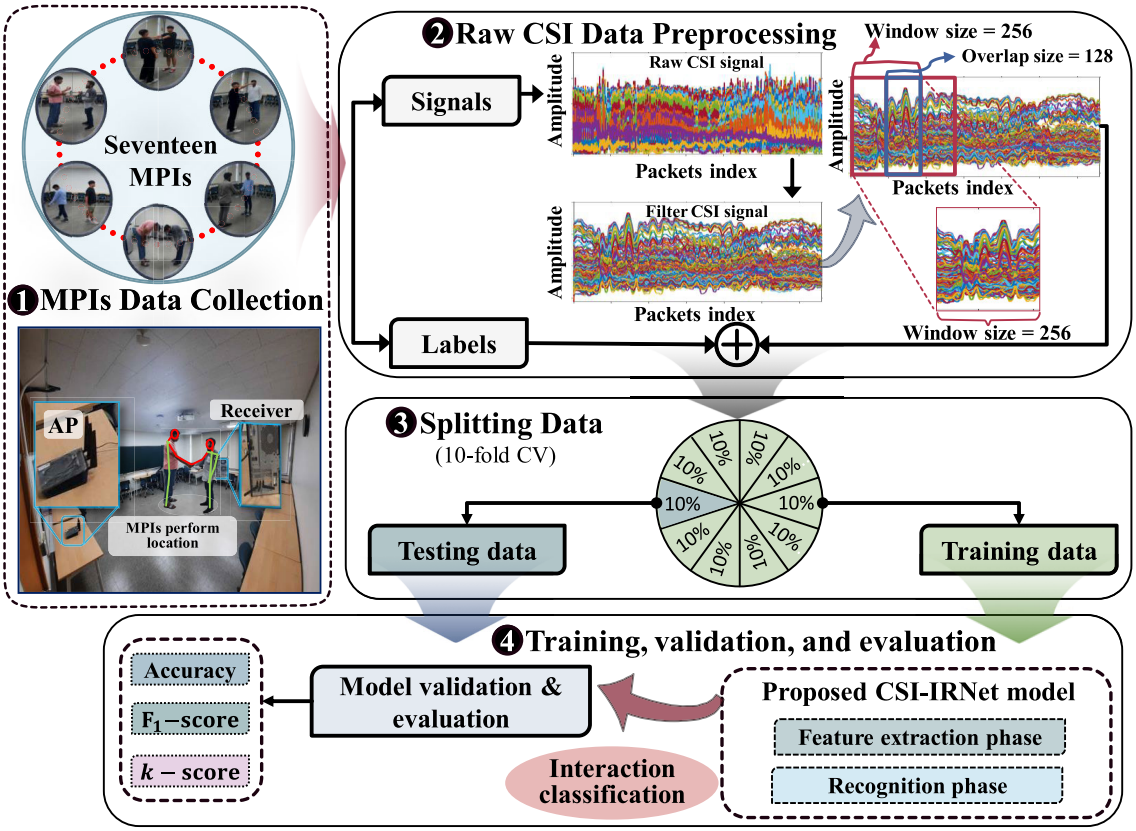


FIGURE 11. Block diagram of the proposed MPI recognition system.

spatial information. Here, $D = N_{\mathrm{p}} \times N_{\mathrm{Sc}}$ where $N_{\mathrm{p}} = N_{\mathrm{Tx}} \times N_{\mathrm{Rx}}$ represents the number of transmit and receive antenna pairs within the testbed. We use a window size of 256 with a 50% overlap, which helps to alleviate concerns related to diverse trial lengths and enhances computational efficiency.

B. PROPOSED CSI-IRNET MODEL

The proposed CSI-IRNet model comprises two primary phases: feature extraction and classification, as illustrated in Fig. 12. The feature extraction phase is responsible for extracting fine-grained features, which are then subsequently utilized by the classification phase for accurate MPI recognition. The feature extraction process starts with a 2D CNN, followed by batch normalization (BN) and ReLU activation, three FSAAM blocks, and six DS-Conv blocks. The DS-Conv block extracts fine-grained features, reducing parameters more efficiently than conventional CNN-based approaches. Meanwhile, the FSAAM block discerns information by identifying regions in the feature maps that require emphasis or suppression. In contrast, the classification phase consists of global average pooling (GAP), dense, and softmax layers. Consequently, the proposed CSI-IRNet model combines DS-Conv and FSAAM to acquire robust feature representations, reducing parameters without compromising the accuracy of MPI recognition. The summary of the CSI-IRNet model is presented in Table 4 and Table 5, respectively.

TABLE 4. Summary for the feature extraction phase of CSI-IRNet model.

Layer name	Output size	Trainable parameters
Conv2D	$128 \times 15 \times 32$	2,624
BN	$128 \times 15 \times 32$	128
DS-Conv	$64 \times 8 \times 64$	2,816
FSAAM	$64 \times 8 \times 64$	4,105
DS-Conv	$32 \times 4 \times 128$	9,152
DS-Conv	$16 \times 2 \times 128$	18,816
FSAAM	$16 \times 2 \times 128$	16,393
DS-Conv	$8 \times 1 \times 256$	35,840
DS-Conv	$4 \times 1 \times 256$	70,400
FSAAM	$4 \times 1 \times 256$	65,545
DS-Conv	$2 \times 1 \times 512$	137,216

TABLE 5. Summary for the classification phase of CSI-IRNet model.

Layer Name	Output Size	Trainable parameters
GAP	1×512	0
Dropout (0.20)	1×512	0
Dense	512×64	32,832
Output	1×18	1,170

1) DEPTHWISE SEPARABLE CONVOLUTION (DS-CONV)

DS-Conv is a convolutional operation that decomposes the traditional convolution into two distinct steps: depthwise convolution (D_{conv}) and pointwise convolution (P_{conv}). D_{conv} is applied to each channel independently and captures features within each channel of the input feature map. After D_{conv}, P_{conv} combines features across channels. This separation enhances computational efficiency and reduces

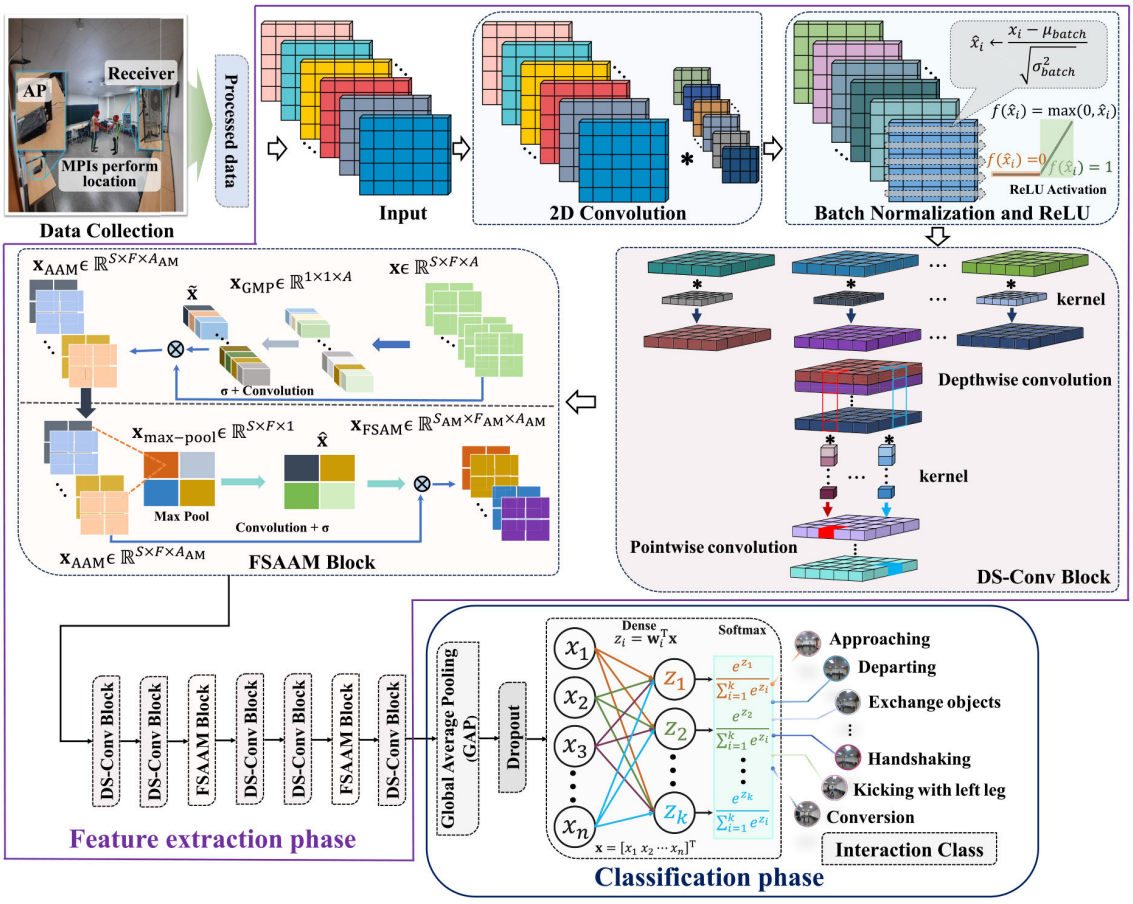


FIGURE 12. Architecture of the CSI-IRNet model.

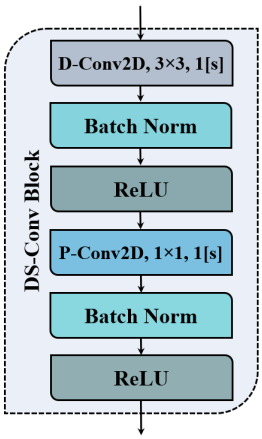


FIGURE 13. DS-Conv block.

the model size, making it particularly advantageous for applications with limited hardware resources.

Let $\mathbf{K}$ and $\mathbf{x}$ represent the convolution kernel and input feature map, respectively. The subscripts i_{h}, i_{w}, and i_{d} denote the input feature map's height, width, and depth, respectively, while k_{h} and k_{w} represent the kernel's height and width, respectively. The depthwise convolution (D_{conv}) with one filter per input channel is expressed as follows:

$$D_{\mathsf{conv}}(\mathbf{K}, \mathbf{x})_{i_{\mathsf{h}}, i_{\mathsf{w}}, i_{\mathsf{d}}} = \sum_{k_{\mathsf{h}}, k_{\mathsf{w}}} \mathbf{K}_{k_{\mathsf{h}}, k_{\mathsf{w}}, i_{\mathsf{d}}} \cdot \mathbf{x}_{i_{\mathsf{h}}+k_{\mathsf{h}}-1, i_{\mathsf{w}}+k_{\mathsf{w}}-1, i_{\mathsf{d}}}. \tag{4}$$

Following D_{conv}, (P_{conv}) is applied:

$$P_{\mathsf{conv}}(\mathbf{K}, \mathbf{x})_{i_{\mathsf{h}}, i_{\mathsf{w}}} = \sum_{i_{\mathsf{d}}} \mathbf{K}_{i_{\mathsf{d}}} \cdot \mathbf{x}_{i_{\mathsf{h}}, i_{\mathsf{w}}, i_{\mathsf{d}}}. \tag{5}$$

Every DS-Conv block is composed of D_{conv} layer with 3×3 size kernels, BN layer, ReLu, and P_{conv} with 1×1 size kernels. Following each D_{conv} and P_{conv} layer, there is a subsequent layer of BN and ReLU which is shown in Fig. 13.

2) FRAME SUBCARRIER ANTENNA ATTENTION MECHANISM (FSAAM)

The movement of various body parts is involved when the pair of humans engage in any kind of MPI. The human body movement patterns are different for each kind of MPI. Wi-Fi signals are affected by the reflection, refraction, and scattering caused by the movement of human body pairs within the space between the transmit and receive antennas as the signals propagate. We focus on extracting features related to CSI signal variations concerning transmit-receive antenna

pairs, alongside exploring the relation of each subcarrier variation with respect to the frames. To facilitate this, we introduce the FSAAM, which encompasses the antenna attention module (AAM) and the frame subcarrier attention module (FSAM). The AAM and FSAM collectively focus on the relationships between antennas, subcarriers, and frames, aiming to derive fine-grained features. The description of the AAM and the FSAM modules is given below.

a: ANTENNA ATTENTION MODULE (AAM)

This module is designed to extract the inter-relationship of different features between the transmit and receive antenna pairs. At the initial step, the AAM applies global max pooling (GMP) to the outcome of the DS-Conv block ($\mathbf{x}$). Then a convolution operation followed by a sigmoid activation function is applied to produce the output ($\widetilde{\mathbf{x}}$). The final step involves an element-wise multiplication between $\widetilde{\mathbf{x}}$ and $\mathbf{x}$ to get the AAM features map $\mathbf{x}_{\mathsf{AAM}}$,

$$\mathbf{x}_{\mathsf{GMP}} = \mathsf{GMP}(\mathbf{x}), \tag{6}$$

$$\widetilde{\mathbf{x}} = \sigma\left(f^{1\times 1}[\mathbf{x}_{\mathsf{GMP}}]\right), \tag{7}$$

$$\mathbf{x}_{\mathsf{AAM}} = \mathbf{x} \otimes \widetilde{\mathbf{x}}, \tag{8}$$

where $\mathbf{x}\in\mathbb{R}^{F\times S\times A}$, $\mathbf{x}_{\mathsf{GMP}} \in \mathbb{R}^{1\times 1\times A}$, and $\mathbf{x}_{\mathsf{AAM}} \in \mathbb{R}^{F\times S\times A_{\mathsf{AM}}}$. F, S, A, σ, and $\otimes$ denote the number of frames, subcarriers, antennas, sigmoid activation function, and element-wise multiplication, respectively.

b: FRAME SUBCARRIER ATTENTION MODULE (FSAM)

The FSAM extracts spatial features by leveraging the relationships between different features across frames and subcarriers. Initially, max pooling applies to the outcomes of the AAM ($\mathbf{x}_{\mathsf{AAM}}$). Subsequently, a convolution operation with a 3×3 filter is performed, followed by a sigmoid activation function to produce $\hat{\mathbf{x}}$. Finally, the FSAM features map $\mathbf{x}_{\mathsf{FSAM}}$ is obtained by applying an element-wise multiplication between $\mathbf{x}_{\mathsf{AAM}}$ and $\hat{\mathbf{x}}$,

$$\mathbf{x}_{\mathsf{max\text{-}pool}} = \mathsf{MaxPool}(\mathbf{x}_{\mathsf{AAM}}), \tag{9}$$

$$\hat{\mathbf{x}} = \sigma\left(f^{3\times 3}[\mathbf{x}_{\mathsf{max\text{-}pool}}]\right), \tag{10}$$

$$\mathbf{x}_{\mathsf{FSAM}} = \mathbf{x}_{\mathsf{AAM}} \otimes \hat{\mathbf{x}}, \tag{11}$$

where $\mathbf{x}_{\mathsf{max\text{-}pool}} \in \mathbb{R}^{F\times S\times 1}$ and $f^{3\times 3}$ denotes a 3×3 kernel.

C. THE EXISTING WI-FI CSI-BASED DL RECOGNITION MODELS

To facilitate a fair comparison, we select two DL models, E2EDLF [16] and CSI-IANet [15]. The E2EDLF [16] model is proposed by the existing Wi-HHI [40] dataset collectors' research group. The E2EDLF [16] model has three phases: input, feature extraction, and recognition. In the input phase, raw CSI signals are converted into CSI images, and the feature extraction phase employs CNN layers to extract features. The recognition phase utilizes these features to recognize the HHIs. Another DL-based model, CSI-IANet [15], first introduces modified inception CNN with a feature attention mechanism to recognize HHIs. The CSI-IANet [15] has three steps: data processing, feature extraction, and recognition. The second-order Butterworth low-pass filter denoises the raw CSI signal, and the feature extraction layer extracts fine-grained features. Finally, the recognition phase classifies HHIs.

D. HYPERPARAMETERS, TRAINING, AND EVALUATION METRICS

A classification model involves three key phases: (i) development, where hyperparameters are selected; (ii) training and validation; and (iii) evaluation. The efficacy of a model's construction and training hinges on the availability of a sufficiently diverse dataset and the appropriate choice of hyperparameters, including learning rate (α), batch size, number of epochs, and activation function. The training set is utilized for selecting hyperparameters, while the validation set is employed for assessing performance. The training process utilizes specific hyperparameters, including α of 1×10^{-3}, 100 epochs, and batch size of 128. A callback monitor is integrated to enhance adaptability, dynamically adjusting α by reducing it by 75% if no improvement is observed over ten consecutive epochs.

In this work, we employ the 10-fold cross-validation (CV) technique that involves random partitioning of the entire dataset into ten non-overlapping subsets of equal size. All the models undergo training through an iterative process across nine folds, with the tenth fold reserved for performance evaluation. The overall recognition performance is determined by averaging the results from all folds. We utilize the Adam optimizer [44] to update weights and the cross-entropy loss function [45] for loss computation.

The assessment of the performance of all the models involves the consideration of three key performance metrics: accuracy, F_1-score, and Cohen's kappa (k-score). These metrics are expressed mathematically in terms of key components, such as true-positive (TP: true positive occurs when the model accurately recognizes the true positive class), false-positive (FP: instances where actual inspection indicates false facts, but experiments identify them as true), true-negative (TN: instances where both actual inspection and experiments indicate actual negative facts), and false-negative (FN: instances where actual inspection indicates true facts, but experiments identify them as false).

$$\mathsf{Recall} = \frac{\mathsf{TP}}{\mathsf{TP}+\mathsf{FN}}, \tag{12}$$

$$\mathsf{Precision} = \frac{\mathsf{TP}}{\mathsf{TP}+\mathsf{FP}}, \tag{13}$$

$$\mathsf{Accuracy} = \frac{\mathsf{TP}+\mathsf{TN}}{\mathsf{TP}+\mathsf{TN}+\mathsf{FP}+\mathsf{FN}}, \tag{14}$$

$$\mathsf{F}_1\text{-score} = 2\times\frac{\mathsf{Recall}\times\mathsf{Precision}}{\mathsf{Recall}+\mathsf{Precision}}. \tag{15}$$

Precision measures the accuracy of predicted true facts among the total actual true facts, while recall gauges how

TABLE 6. Results obtained from the 10-fold CV of the CSI-IRNet model on the developed Wi-MIR dataset.

Metrics	Fold										Average
	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	
Accuracy (%)	94.10	94.03	94.04	94.06	94.05	93.72	93.83	94.15	94.26	93.91	94.02 ± 0.15
F_1-score (%)	93.25	93.16	93.17	93.21	93.20	92.75	92.92	93.34	93.48	92.94	93.14 ± 0.20
k-score	0.937	0.936	0.938	0.937	0.937	0.935	0.934	0.938	0.939	0.935	0.937 ± 0.002

TABLE 7. Results obtained from the ML classifiers on the developed Wi-MIR dataset.

Dataset	Classifier	Metrics (%)	
		Accuracy	F_1-score
Wi-MIR	RF [11]	65.52	64.10
	SVM [11]	62.06	60.12
	k-NN [11]	60.00	57.00
	DT [16]	42.09	38.21
	CSI-IRNet (proposed)	94.02	93.12

often the model correctly identifies instances based on the true positive rate. The F_1-score, a harmonic mean of recall and precision, offers a more insightful performance measure than accuracy, whereas k-score indicates the classifier's performance by making random frequency estimation based on the number of samples for each class, yielding values between 0 and 1.

VII. RESULT AND DISCUSSION

This section presents the findings derived in the case of the developed Wi-MIR and the existing Wi-HHI [40] dataset, in terms of MPI recognition accuracy, F_1-score, and k-score. Furthermore, the robustness and consistency of the evaluation metrics (accuracy, F_1-score, and k-score) affirm the appropriateness of the developed dataset's representation and features for modeling MPIs. We conduct a performance comparison between the developed Wi-MIR and the existing Wi-HHI [40] datasets by applying the ML and the existing DL-based models along with the proposed attention-based CSI-IRNet model.

A. EXPERIMENTAL RESULTS ON DEVELOPED WI-MIR DATASET

1) PERFORMANCE OF THE PROPOSED CSI-IRNET MODEL

To validate the developed Wi-MIR dataset for classifying MPIs, we perform an evaluation using the proposed CSI-IRNet model. This experiment aims to establish the reliability and trustworthiness of the developed dataset in accurately recognizing MPIs. The comprehensive evaluation results of the proposed CSI-IRNet model on the developed Wi-MIR dataset, utilizing the 10-fold CV approach, are shown in Table 6. The tabulated results in Table 6 provide the model's performance metrics across each fold and their corresponding averages. The accuracy, F_1-score, and k-score are achieved with an average and standard deviation of $94.02\% \pm 0.15\%$, $93.14\% \pm 0.20\%$, and 0.937 ± 0.002, respectively. Noteworthy, the observations from the detailed analysis of Table 6 reveal that among the ten folds, the 8-th fold attains the maximum accuracy performance of 94.15%, while the remaining folds consistently achieve accuracy levels surpassing 93.72%.

In addition, Fig. 14 illustrates the confusion matrix generated by the proposed CSI-IRNet model, providing valuable insights into its recognition performance across seventeen MPIs. The main diagonal represents the average recognition accuracy for each class, while off-diagonal elements indicate misclassifications. From a closer observation of the confusion matrix, the proposed CSI-IRNet model achieves high accuracy in recognizing "Touching another person" of 98.37%, and all the other MPIs achieve more than 93% accuracy except steady state, Approaching, and Departing MPIs.

2) PERFORMANCE OF DIFFERENT ML CLASSIFIERS

A comprehensive overview of the performance of four ML classifiers applying to the developed Wi-MIR dataset for MPI recognition is tabulated in Table 7. The classifiers considered for evaluation include RF, SVM, k-nearest neighbors (k-NN), and decision tree (DT). The RF classifier outperforms others, achieving an accuracy of 65.52% and F_1-score of 64.10%. Following closely, the SVM demonstrates competitive results with an accuracy of 62.06% and F_1-score of 60.12%. The k-NN and DT classifiers exhibit accuracy of 60% and 42.09%, respectively, with corresponding F_1-score of 57% and 38.21%. The above performance results show that the RF and SVM have better recognition outcomes than the other two (k-NN and DT) classifiers. In contrast, the proposed CSI-IRNet model achieves an impressive accuracy and F_1-score of 94.02% and 93.12%, respectively.

B. COMPARISON BETWEEN DEVELOPED WI-MIR AND THE EXISTING WI-HHI DATASET

We compare the developed Wi-MIR dataset with only one publicly available Wi-HHI [40] dataset to evaluate how well our dataset captures the intricacies of MPIs, like the richness of features and specific characteristics of MPIs. The existing Wi-HHI [40] dataset (refer to Section II) has 12 different HHIs with one steady state, while our dataset contains seventeen MPIs with a steady state. Table 8 presents a qualitative comparison between the developed Wi-MIR and the existing Wi-HHI [40] datasets. Since the Wi-HHI [40] dataset involves 12 MPIs, we deliberately isolate a subset of 12 identical MPIs from the developed Wi-MIR dataset, aligning with the Wi-HHI [40] dataset for implementing the DL models. The details of the identical MPIs and the average number of recorded Wi-Fi packets for each MPI of both

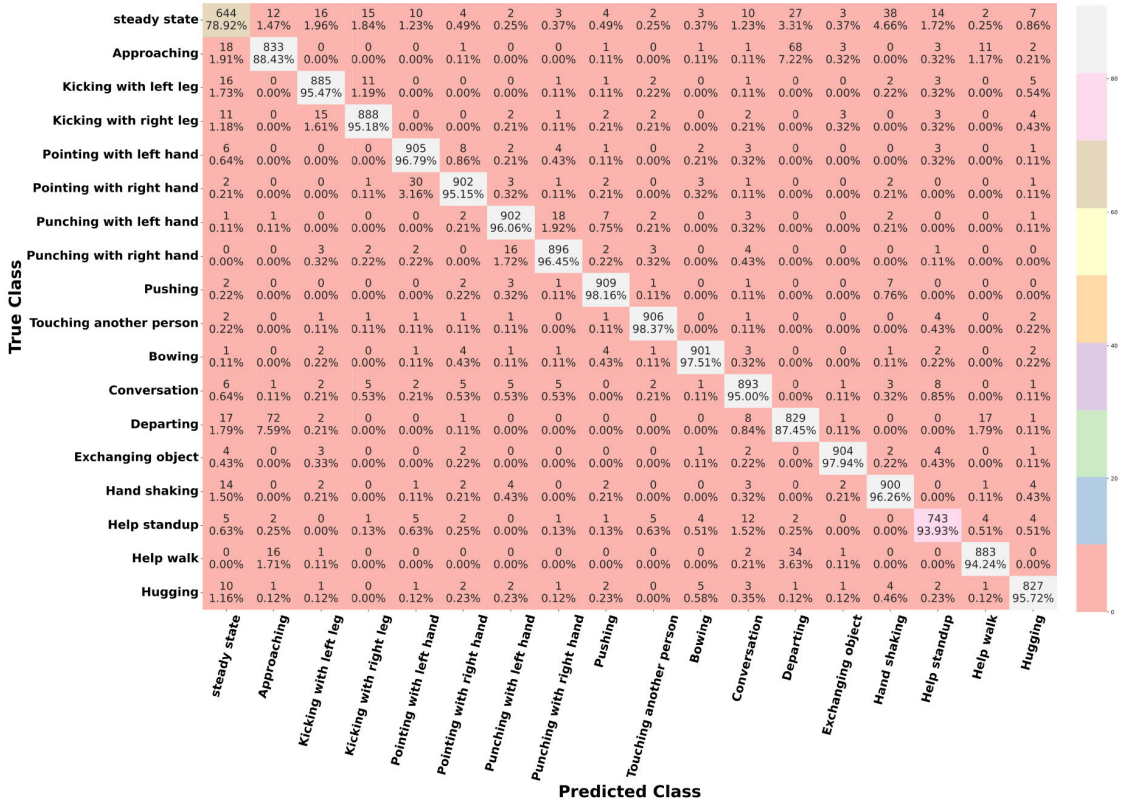


FIGURE 14. Confusion matrix of the proposed model for MPI recognition on the developed Wi-MIR dataset.

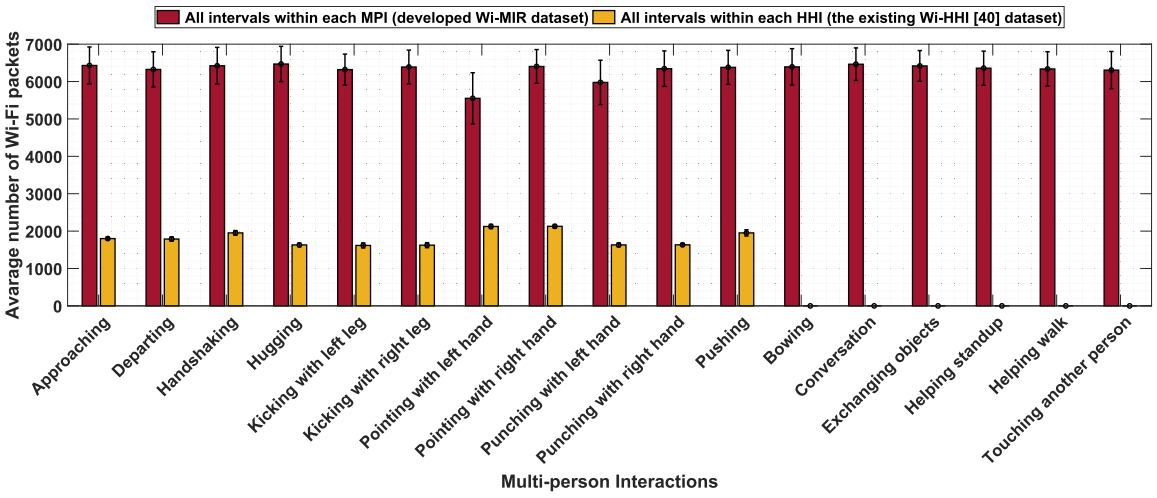


FIGURE 15. Comparison of average Wi-Fi packets for each MPI on developed Wi-MIR and the existing Wi-HHI [40] dataset.

datasets (developed Wi-MIR and the existing Wi-HHI [40]) are shown in Fig. 15.

1) MPI RECOGNITION PERFORMANCE COMPARISON

We select two existing DL models: E2EDLF [16] and CSI-IANet [15] (refer to Section VI-C) to facilitate a fair comparison with our proposed CSI-IRNet. Table 9 provides an in-depth comparative performance analysis of the DL models (E2EDLF [16], CSI-IANet [15], and CSI-IRNet) applied to the existing Wi-HHI [40] and developed Wi-MIR dataset in the case of 12 and 18 MPIs, respectively. For considering the existing Wi-HHI [40] dataset for 12 MPI recognition, the E2EDLF [16] model achieves an accuracy, F_1-score, and k-score of 86.30%, 86%, and 0.85, respectively. The CSI-IANet [15] exhibits improved performance, reaching an accuracy of 91.30%, F_1-score of 91.27%, and a k-sore of 0.89. The proposed CSI-IRNet outperforms both models with an accuracy of 94.55%, F_1-score of 94.50%, and

TABLE 8. Comparison of the existing Wi-HHI [40] and developed Wi-MIR datasets.

Aspect	Wi-HHI [40]	Developed Wi-MPI
Hardware	Sagemcom 2,704, 2.4 GHz, and 20 MHz	NETGEAR Nighthawk-R7000 router router, 5 GHz, and 20 MHz
MIMO	2×3 MIMO ($N_{\text{Tx}}=2$, $N_{\text{Rx}}=3$)	3×3 MIMO ($N_{\text{Tx}}=3$, $N_{\text{Rx}}=3$)
CSI Tool	CSI tool [42]	CSI tool [42]
Subcarrier	6×30 subcarriers	9×30 subcarriers
Sampling Rate	$\approx$ 330 Hz	$\approx$ 950 Hz
MPIs	12 interactions	17 interactions
Pairs/Trials	40 pairs, 10 trials	11 pairs, 20 trials
Testbed size	5.3m $\times$ 5.3m	8m $\times$ 6.1m
Remarks	Publicly available, valuable insights into HHI. Operates at a lower sampling rate and 2×3 MIMO. Consider resource ML and DL applications.	Publicly available, potential advantages with higher sampling rate and 3×3 MIMO. Covering more diverse kinds of MPIs. A promising resource for ML and DL applications.

TABLE 9. Comparing results of existing and proposed DL models on the existing Wi-HHI [40] and Wi-MIR datasets.

DL model	Accuracy (%)	F_1-score (%)	k-score
	12 class (the existing Wi-HHI [40])		
E2EDLF [16]	86.30	86.00	0.85
CSI-IANet [15]	91.30	91.27	0.89
CSI-IRNet (proposed)	**94.55**	**94.50**	**0.94**
	12 class (developed Wi-MIR)		
E2EDLF [16]	87.59	87.12	0.86
CSI-IANet [15]	93.49	93.4	0.92
CSI-IRNet (proposed)	**96.18**	**95.15**	**0.95**
	18 class (developed Wi-MIR)		
E2EDLF [16]	84.46	82.05	0.82
CSI-IANet [15]	92.46	91.59	0.91
CSI-IRNet (proposed)	**94.02**	**93.14**	**0.94**

k-score of 0.94. To compare the performance of the developed Wi-MIR with the existing Wi-HHI [40] dataset, we separate 12 identical MPIs among 18 available MPIs and apply the DL models in MPI recognitions. The E2EDLF [16] model achieves an accuracy of 87.59%, F_1-score of 87.12%, and k-score of 0.86. The CSI-IANet [15] achieves an accuracy of 93.49%, F_1-score of 93.40%, and k-score of 0.92. The proposed CSI-IRNet shows better performance, achieving an accuracy of 96.18%, F_1-score of 95.15%, and k-score of 0.95. All the DL (the existing and proposed) models individually get better MPI recognition performance on the developed Wi-MIR dataset than the existing Wi-HHI [40] dataset. The proposed CSI-IRNet outperforms the other two existing DL models in terms of the accuracy, F_1-score, and k-score for both the Wi-HHI [40] and Wi-MIR dataset.

Expanding the evaluation to the developed Wi-MIR dataset with 18 MPIs, all DL (the existing and proposed) models exhibit consistent performance trends. The E2EDLF [16] model achieves an accuracy of 84.46%, F_1-score of 82.05%, and k-score of 0.82. The CSI-IANet [15] demonstrates improved accuracy at 92.46%, F_1-score of 91.59%, and k-score of 0.911. Our proposed CSI-IRNet maintains its better performance, achieving an accuracy of 94.02%, F_1-score of 93.14%, and k-score of 0.937. These results underscore the efficacy of CSI-IRNet in accurately recognizing 18 MPIs. The performance of the proposed CSI-IRNet model, which outperforms the two DL models (E2EDLF [16] and CSI-IANet [15]), can be attributed to two key factors: i) the integration of the DS-Conv and the FSAAM block enhances the model's capacity to concentrate on the most pertinent features while disregarding irrelevant ones, and ii) the inclusion of extra layers, like dropout, and a well-chosen number of dense layers. These two factors collectively substantiate the proposed CSI-IRNet model's efficacy in handling complex CSI signals for accurate MPI recognition. However, the overall competent performance of the models on the developed Wi-MIR dataset with 18 class MPIs shows a slight decrease in accuracy compared to the 12 class MPIs. This decline in performance can be attributed to the increased number of MPI classes that introduce additional complexity and challenges.

From the performance comparison result illustrated in Table 9, it is obvious that there is a remarkable improvement in MPI recognition performance in the case of the developed Wi-MIR than the existing Wi-HHI [40] dataset. Operating at a 5 GHz band with a 20 MHz channel bandwidth and utilizing a 3×3 MIMO system, the developed Wi-MIR dataset captures richer information, including 270 ($3 \times 3 \times 30$) subcarriers. With a higher sampling rate of 950 Hz, this dataset enhances the granularity of captured data, providing a robust foundation for training and inference in MPI recognition models. The expanded dataset size and the higher sampling rate enhance the model's ability to discern subtle patterns and variations in the Wi-Fi signals associated with different MPIs. The thorough evaluation of the developed Wi-MIR dataset demonstrates its reliability, affirming its suitability for online availability and facilitating its utilization by other researchers.

VIII. CONCLUSION

In this paper, we have developed and publicly released a Wi-Fi CSI-based MPI recognition dataset, coined *Wi-MIR*, that encompasses seventeen distinct MPIs and collected by eleven human pairs in an indoor environment. This dataset utilizes more subcarriers and a higher sampling rate to curate a collection of seventeen MPIs, including five new kinds of MPIs. Moreover, the representation and characteristics of the developed Wi-MIR have been compared with the existing Wi-HHI [40] dataset and evaluated by the ML classifiers, the existing DL-based models, and the proposed CSI-IRNet models in terms of the accuracy, F_1-score, and k-score. According to performance comparison, the

developed Wi-MIR dataset outperforms the existing dataset across all evaluation metrics, establishing it as a valuable resource for MPI research. Future work aims to enhance CSI data gathering by exploring newer Wi-Fi APs that support the IEEE 802.11ax and 802.11bf standards, which offer more subcarriers and improved signal quality. By doing so, it enables greater accuracy and diversity in data collection. Additionally, we plan to broaden our dataset to cover a wide range of daily life MPIs, enrich indoor environments with NLOS, and account for diverse age and gender dynamics.

REFERENCES

[1] X. Chen, Y. Zou, C. Li, and W. Xiao, "A deep learning based lightweight human activity recognition system using reconstructed WiFi CSI," *IEEE Trans. Human-Mach. Syst.*, vol. 54, no. 1, pp. 68–78, Feb. 2024.

[2] A. Hussain, S. U. Khan, N. Khan, I. Rida, M. Alharbi, and S. W. Baik, "Low-light aware framework for human activity recognition via optimized dual stream parallel network," *Alexandria Eng. J.*, vol. 74, pp. 569–583, Jul. 2023.

[3] X. Zhou, W. Liang, K. I.-K. Wang, H. Wang, L. T. Yang, and Q. Jin, "Deep-learning-enhanced human activity recognition for Internet of Healthcare Things," *IEEE Internet Things J.*, vol. 7, no. 7, pp. 6429–6438, Jul. 2020.

[4] Z. Hussain, M. Sheng, and W. E. Zhang, "Different approaches for human activity recognition: A survey," Jun. 2019, *arXiv:1906.05074*.

[5] Y.-L. Hsu, S.-C. Yang, H.-C. Chang, and H.-C. Lai, "Human daily and sport activity recognition using a wearable inertial sensor network," *IEEE Access*, vol. 6, pp. 31715–31728, 2018.

[6] O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 1192–1209, Nov. 2012.

[7] H. Zhang and L. E. Parker, "CoDe4D: Color-depth local spatio-temporal features for human activity recognition from RGB-D videos," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 3, pp. 541–555, Mar. 2016.

[8] C. Zhu and W. Sheng, "Wearable sensor-based hand gesture and daily activity recognition for robot-assisted living," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 41, no. 3, pp. 569–573, May 2011.

[9] I. Ullmann, R. G. Guendel, N. C. Kruse, F. Fioranelli, and A. Yarovoy, "A survey on radar-based continuous human activity recognition," *IEEE J. Microw.*, vol. 3, no. 3, pp. 938–950, Jul. 2023.

[10] M. S. Islam, M. K. A. Jannat, M. N. Hossain, W.-S. Kim, S.-W. Lee, and S.-H. Yang, "STC-NLSTMNet: An improved human activity recognition method using convolutional neural network with NLSTM from WiFi CSI," *Sensors*, vol. 23, no. 1, p. 356, Dec. 2022.

[11] M. K. A. Jannat, M. S. Islam, S.-H. Yang, and H. Liu, "Efficient Wi-Fi-based human activity recognition using adaptive antenna elimination," *IEEE Access*, vol. 11, pp. 105440–105454, 2023.

[12] M. H. Kabir, Md. A. Hasan, and W. Shin, "CSI-DeepNet: A lightweight deep convolutional neural network based hand gesture recognition system using Wi-Fi CSI signal," *IEEE Access*, vol. 10, pp. 114787–114801, 2022.

[13] C. Feng, S. Arshad, S. Zhou, D. Cao, and Y. Liu, "Wi-Multi: A three-phase system for multiple human activity recognition with commercial WiFi devices," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 7293–7304, Aug. 2019.

[14] W. Jiao and C. Zhang, "An efficient human activity recognition system using WiFi channel state information," *IEEE Syst. J.*, vol. 17, no. 4, pp. 6687–6690, Dec. 2023.

[15] M. H. Kabir, M. H. Rahman, and W. Shin, "CSI-IANet: An inception attention network for human-human interaction recognition based on CSI signal," *IEEE Access*, vol. 9, pp. 166624–166638, 2021.

[16] R. Alazrai, M. Hababeh, B. A. Alsaify, M. Z. Ali, and M. I. Daoud, "An end-to-end deep learning framework for recognizing human-to-human interactions using Wi-Fi signals," *IEEE Access*, vol. 8, pp. 197695–197710, 2020.

[17] M. Humayun Kabir, K. Hyun-Wook, and Y. Sung-Hyun, "The development of a modular type platform for home network service," in *Proc. Int. Conf. Informat., Electron. Vis. (ICIEV)*, May 2012, pp. 582–587.

[18] Z. Chen, L. Zhang, C. Jiang, Z. Cao, and W. Cui, "WiFi CSI based passive human activity recognition using attention based BLSTM," *IEEE Trans. Mobile Comput.*, vol. 18, no. 11, pp. 2714–2724, Nov. 2019.

[19] L. Guo, L. Wang, C. Lin, J. Liu, B. Lu, J. Fang, Z. Liu, Z. Shan, J. Yang, and S. Guo, "Wiar: A public dataset for WiFi-based activity recognition," *IEEE Access*, vol. 7, pp. 154935–154945, 2019.

[20] S. Yousefi, H. Narui, S. Dayal, S. Ermon, and S. Valaee, "A survey on behavior recognition using WiFi channel state information," *IEEE Commun. Mag.*, vol. 55, no. 10, pp. 98–104, Oct. 2017.

[21] A. Manzi, L. Fiorini, R. Limosani, P. Dario, and F. Cavallo, "Two-person activity recognition using skeleton data," *IET Comput. Vis.*, vol. 12, no. 1, pp. 27–35, Feb. 2018.

[22] A. Stergiou and R. Poppe, "Analyzing human–human interactions: A survey," *Comput. Vis. Image Understand.*, vol. 188, Nov. 2019, Art. no. 102799.

[23] B. Wei, W. Hu, M. Yang, and C. T. Chou, "Radio-based device-free activity recognition with radio frequency interference," in *Proc. 14th Int. Conf. Inf. Process. Sensor Netw.*, Apr. 2015, pp. 154–165.

[24] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and modeling of WiFi signal based human activity recognition," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, Sep. 2015, pp. 65–76.

[25] X. Zheng, J. Wang, L. Shangguan, Z. Zhou, and Y. Liu, "Smokey: Ubiquitous smoking detection with commercial WiFi infrastructures," in *Proc. 35th Annu. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, Apr. 2016, pp. 1–9.

[26] W. Jiang, C. Miao, F. Ma, S. Yao, Y. Wang, Y. Yuan, H. Xue, C. Song, X. Ma, D. Koutsonikolas, W. Xu, and L. Su, "Towards environment independent device free human activity recognition," in *Proc. 24th Annu. Int. Conf. Mobile Comput. Netw.*, Oct. 2018, pp. 289–304.

[27] J. Schäfer, B. R. Barrsiwal, M. Kokhkharova, H. Adil, and J. Liebehenschel, "Human activity recognition using CSI information with nexmon," *Appl. Sci.*, vol. 11, no. 19, p. 8860, Sep. 2021.

[28] Y. Wang, K. Wu, and L. M. Ni, "WiFall: Device-free fall detection by wireless networks," *IEEE Trans. Mobile Comput.*, vol. 16, no. 2, pp. 581–594, Feb. 2016.

[29] F. Zhang, K. Niu, J. Xiong, B. Jin, T. Gu, Y. Jiang, and D. Zhang, "Towards a diffraction-based sensing approach on human activity recognition," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 3, no. 1, pp. 1–25, Mar. 2019.

[30] F. Wang, J. Feng, Y. Zhao, X. Zhang, S. Zhang, and J. Han, "Joint activity recognition and indoor localization with WiFi fingerprints," *IEEE Access*, vol. 7, pp. 80058–80068, 2019.

[31] B. A. Alsaify, M. M. Almazari, R. Alazrai, and M. I. Daoud, "A dataset for Wi-Fi-based human activity recognition in line-of-sight and non-line-of-sight indoor environments," *Data Brief*, vol. 33, Dec. 2020, Art. no. 106534.

[32] F. Meneghello, N. Dal Fabbro, D. Garlisi, I. Tinnirello, and M. Rossi, "A CSI dataset for wireless human sensing on 80 MHz Wi-Fi channels," *IEEE Commun. Mag.*, vol. 61, no. 9, pp. 146–152, Jun. 2023.

[33] I. Galdino, J. C. H. Soto, E. Caballero, V. Ferreira, T. C. Ramos, C. Albuquerque, and D. C. Muchaluat-Saade, "EHealth CSI: A Wi-Fi CSI dataset of human activities," *IEEE Access*, vol. 11, pp. 71003–71012, 2023.

[34] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Device-free human activity recognition using commercial WiFi devices," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 5, pp. 1118–1131, May 2017.

[35] H. Salehinejad and S. Valaee, "LiteHAR: Lightweight human activity recognition from WiFi signals with random convolution kernels," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2022, pp. 4068–4072.

[36] S. K. Yadav, S. Sai, A. Gundewar, H. Rathore, K. Tiwari, H. M. Pandey, and M. Mathur, "CSITime: Privacy-preserving human activity recognition using WiFi channel state information," *Neural Netw.*, vol. 146, pp. 11–21, Feb. 2022.

[37] M. Abdel-Basset, H. Hawash, N. Moustafa, and N. Mohammad, "H2HI-net: A dual-branch network for recognizing human-to-human interactions from channel-state information," *IEEE Internet Things J.*, vol. 9, no. 12, pp. 10010–10021, Jun. 2022.

[38] B. A. Alsaify, M. M. Almazari, R. Alazrai, S. Alouneh, and M. I. Daoud, "A CSI-based multi-environment human activity recognition framework," *Appl. Sci.*, vol. 12, no. 2, p. 930, Jan. 2022.

[39] X. Yang, D. Zhai, R. Zhang, H. Cao, S. Garg, and M. M. Hassan, "Human-to-human interaction behaviors sensing based on complex-valued neural network using Wi-Fi channel state information," *Future Gener. Comput. Syst.*, vol. 148, pp. 160–172, Nov. 2023.

[40] R. Alazrai, A. Awad, B. Alsaify, M. Hababeh, and M. I. Daoud, "A dataset for Wi-Fi-based human-to-human interaction recognition," *Data Brief*, vol. 31, Aug. 2020, Art. no. 105668.
[41] Shin Lab @ Korea Univ. (Jan. 10, 2024). *Wi-MIR: A CSI Dataset for Wi-Fi Based Multi-Person Interaction Recognition*. [Online]. Available: https://sites.google.com/view/multi-person-interaction
[42] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11n traces with channel state information," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 1, p. 53, Jan. 2011.
[43] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.
[44] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[45] K. Janocha and W. M. Czarnecki, "On loss functions for deep neural networks in classification," 2017, *arXiv:1702.05659*.

MD. SHAFIQUL ISLAM received the B.Sc. and M.Sc. degrees in computer science and engineering from Islamic University, Kushtia, Bangladesh, in 2016 and 2017, respectively, and the Ph.D. degree from the Department of Electronic Engineering, Kwangwoon University, Seoul, Republic of Korea, in 2023. He is currently a Postdoctoral Researcher with Korea University, Seoul. Prior to joining Korea University, he was a Postdoctoral Researcher with Ajou University, Suwon-si, South Korea. His main research interests include human activity recognition, digital signal and image processing, and deep learning.

M. HUMAYUN KABIR (Member, IEEE) received the B.Sc. and M.Sc. degrees from Islamic University, Kushtia, Bangladesh, in 2001 and 2003, respectively, and the Ph.D. degree from the Department of Electronic Engineering, Kwangwoon University, Seoul, Republic of Korea, in 2016. He was a Postdoctoral Researcher with Ajou University, Seoul, from 2021 to 2023. He is currently a Professor with Islamic University. His research interests include the IoT, satellite communication, ML, and deep-learning-based signal processing. He was awarded the Best Paper at the 32nd Joint Conference on Communication and Information (JCCI) and the Best Workshop Paper at the International Conference on ICT Convergence (ICTC), in 2022. He is acting as a Reviewer of several reputed journals, especially IEEE TRANSACTIONS ON EMERGING TOPICS IN COMPUTATIONAL INTELLIGENCE, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE ACCESS, and *IET Networks*.

MD. ALI HASAN received the B.Sc. degree in electrical and electronic engineering from Islamic University, Kushtia, Bangladesh, in 2020. He is currently pursuing the M.S. degree with Ajou University, Suwon-si, South Korea. His main research interests include machine learning, deep learning-based signal processing, and satellite-based positioning.

WONJAE SHIN (Senior Member, IEEE) received the B.S. and M.S. degrees from Korea Advanced Institute of Science and Technology, in 2005 and 2007, respectively, and the Ph.D. degree from the Department of Electrical and Computer Engineering, Seoul National University (SNU), South Korea, in 2017.

From 2007 to 2014, he was a member of the Technical Staff with the Samsung Advanced Institute of Technology and Samsung Electronics Company Ltd., South Korea, where he contributed to next-generation wireless communication networks, especially for 3GPP LTE/LTE-advanced standardizations. He was a Visiting Scholar and a Postdoctoral Research Fellow with Princeton University, Princeton, NJ, USA, from 2016 to 2018. Since 2023, he has been with the School of Electrical Engineering, Korea University, Seoul, South Korea, where he is currently an Associate Professor. Prior to joining Korea University, he was a Faculty Member with Pusan National University, Busan, and Ajou University, Suwon-si, South Korea. His research interests include the design and analysis of future wireless communication systems, such as interference-limited networks and machine learning for wireless networks. He was a recipient of the Fred W. Ellersick Prize and the Asia-Pacific Outstanding Young Researcher Award from the IEEE Communications Society, in 2020, the International Conference on ICT Convergence (ICTC) Best Workshop Paper Award, in 2022, the *Journal of Korean Institute of Communications and Information Sciences* (J-KICS) Best Paper Award, in 2021, the Best Ph.D. Dissertation Award from SNU, in 2017, the Gold Prize from the IEEE Student Paper Contest (Seoul Section), in 2014, and the Award of the Ministry of Science and ICT of Korea in the IDIS-Electronic News ICT Paper Contest, in 2017. He was a co-recipient of the SAIT Patent Award, in 2010, the Samsung Journal of Innovative Technology Award, in 2010, the Samsung Human Tech Paper Contest, in 2010, and the Samsung CEO Award, in 2013. He was also awarded several fellowships, including the Samsung Fellowship Program, in 2014, and the SNU Long-Term Overseas Study Scholarship, in 2016. He was recognized as an Exemplary Reviewer of IEEE WIRELESS COMMUNICATIONS LETTERS, in 2014, and IEEE TRANSACTIONS ON COMMUNICATIONS, in 2019. He is an Associate Editor of IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY.