IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 1, 2019, accepted October 19, 2019, date of publication October 23, 2019, date of current version November 6, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2949123

A Survey on Human Behavior Recognition Using Channel State Information

ZHENGJIE WANG, KANGKANG JIANG, YUSHAN HOU, WENWEN DOU, CHENGMING ZHANG, ZEHUA HUANG, AND YINJING GUO

College of Electronic and Information Engineering, Shandong University of Science and Technology, Qingdao 266590, China

Corresponding authors: Zhengjie Wang (cieewangzj@163.com) and Yinjing Guo (gyjlwh@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61471224, in part by the Qingdao Postdoctoral Applied Research Project under Grant 2015180, and in part by the Shandong Province Key Research and Development Plan (Public Welfare Special) Project under Grant 2018GHY115022.

ABSTRACT Recently, device-free human behavior recognition has become a hot research topic and has achieved significant progress in the field of ubiquitous computing. Among various implementation, behavior recognition based on WiFi CSI (channel state information) has drawn wide attention due to its major advantages. This paper investigates more than 100 latest CSI based behavior recognition applications within the last 6 years and presents a comprehensive survey from every aspect of human behavior recognition. Firstly, this paper reviews general behavior recognition applications using the WiFi signal and presents the basic concept of CSI and the fundamental principle of CSI-based behavior recognition. This paper analyzes the key components and core characteristics of the system architecture of human behavior recognition using CSI. Afterward, we divide the sensing procedures into many steps and summarize the typical studies from these steps, including base signal selection, signal preprocessing, and identification approaches. Next, based on the recognition technique, we classify the applications into three groups, including pattern-based, model-based, and deep learning-based approach. In every group, we categorize the state-of-the-art applications into three groups, including coarse-grained specific behavior recognition, fine-grained specific behavior recognition, and activity inference. It elaborates the typical behavior recognition applications from five aspects, including experimental equipment, experimental scenario, behavior, classifier, and system performance. Then, this paper presents comprehensive discussions about representative applications from the implementation view and outlines the major consideration when developing a recognition system. Finally, this article concludes by analyzing the open issues of CSI-based behavior recognition applications and pointing out future research directions.

INDEX TERMS Channel state information (CSI), deep learning, human behavior recognition, model, pattern, WiFi.

I. INTRODUCTION

Recent years have witnessed the rapid development of the Internet of Things (IoT). The demand for pervasive computing is dramatically increasing because IoT provides us with useful information about monitoring target and facilities development of ubiquitous computing applications. The IoT technology can be employed on environment monitoring [1], smart city [2], and industrial fields. Besides, this technology also has spread into our daily lives and can be used to sense human behavior [3], such as health monitoring [4], behavior understanding [5], user profile construction [6], activity inference [7], smart home control [8], human localization [9], human position tracking [10], [11], and occupancy detection [12], etc. These applications of human behavior sensing largely enrich the research content and provide us with novel human-computer interaction methods. Traditional human sensing techniques usually require the user to wear or carry some sensors, which restricts their application environments and brings much inconvenience to users [13]. With the widespread deployment of WiFi devices in indoor environments, the device-free WiFi-based behavior recognition technique has drawn more attention because it overcomes some existing shortcomings of common behavior recognition

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Wang.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see http://creativecommons.org/licenses/by/4.0/

system, including the deployment cost, privacy violation, and applicable environment limitations, etc. In other words, this approach can be widely deployed in various environments and work within the Non-Line of Sight (NLOS) path even across the walls.

Currently, WiFi-based activity recognition techniques usually include two types of wireless signal, RSS (radio signal strength) and CSI (channel state information). RSS provides coarse-grained information about communication links while CSI describes fine-grained information about the communication channel state [14]. We can measure RSS using most wireless devices easily because the RSS collection is supposed by almost all wireless chips. As for CSI, we have to modify NIC driver to measure the CSI using COTS devices, including Intel 5300 NIC [15], Atheros AR9382 [16], Atheros AR9462 and AR9480 [17], Atheros 9580 [18], [19], and Atheros 9390 [20], [21]. Alternatively, we can measure CSI using a special device such as SDR [22] or WARP [23], [24]. At present, there has a great many state-of-the-art WiFi-based recognition applications and studies. This part reviews some typical recognition applications based on RSS or CSI.

RSS is a measurement of the power of the radio signal at the received end. In other words, it describes the average signal power of the received signal. Currently, RSS-based approaches have been extensively used in the field of identification and have achieved great recognition performance. In this section, we introduce some typical RSS-based recognition applications, such as object tracking and location [25]–[28], activity monitoring or recognition [29], hand gesture recognition [30], driving behavior identification [31], and crowd counting [32].

Youssef *et al.* [25] first propose the concept of device-free passive (DfP) localization with RSS and present a passive radio map construction and tracking application. In addition, reference [25] displays promising pictures of passive device-free localization and tracking. Wilson and Patwari [26] propose an approach for tracking and localizing the target behind the wall. It adopts a statistical model to depict change rules of the signal by using the variance of RSS caused by user motion. Xiang *et al.* [27] propose an approach to enhance the traditional RSS-based indoor localization precision by using features of light, temperature, and humidity information. Wilson and Patwari [28] propose radio tomographic imaging (RTI) technology for imaging the attenuation. It utilizes a line model for depicting the RSS to acquire images of moving objects and locates a certain person in the area of RF coverage by imaging the attention. Wang *et al*. [29] propose a device-free localization and activity recognition (DFLAR) system using deep learning. This system adopts 3-layer networks to extract features. Compared with traditional handcraft features extraction approaches, deep learning-based methods improve 10% identification precision. Abudulaziz *et al*. propose Wisture [30], an RSS-based hand gestures recognition system on the smartphone. This system can recognize non-contact dynamic hand gestures (e.g., swipe, push, and pull) using Long Short-Term Memory (LSTM) and Recurrent Neural Network (RNN), and it achieves 94% accuracy in gestures detection and classification. Depatla *et al*. propose an RSS-based crowd counting system in through-the-wall scenario [32]. This system tests 20 people in 5 different through-the-wall scenarios (e.g., concrete, plaster, and wood) and achieves the error of two people.

CSI can be expressed as a complex matrix, each entry of which records the amplitude and phase response of the signal transmission channel of each subcarrier. We can compute the amplitude and phase of the received signal of each channel from the complex number and evaluate the channel quality using each entry of the matrix. Therefore, the amplitude of the CSI signal quantifies the signal power attenuation after the multi-path effect, similar to the received signal strength of the wireless signal. Currently, there has many CSI-based recognition applications, such as indoor human behavior recognition, indoor localization [33]–[41], indoor object state detection [42], fire detection [43], traffic monitoring [44], wheat moisture detection [45], object distinction [46], in-baggage suspicious object detection [47], school violence monitoring [48], and other identification applications in driving scenario (e.g., risky driving behavior detection [49], [50], driver fatigue detection [51], driver's distracted behavior detection [52], in-vehicle behavior and hand gesture recognition [53], [54]).

This article concentrates on human behavior recognition with CSI. We present the typical human behavior in this paper to clarify the research area. Based on the behavior recognition purpose, we divide CSI-based indoor human behavior recognition applications into two groups, such as specific behavior recognition and activity inference. Specific behavior recognition refers to identifying some simple human behaviors that we just consider as distinct actions, such as daily behaviors (e.g., walking, running, sitting, cooking, eating, lying, standing, brushing, opening and closing door, etc.), abnormal behaviors (e.g., slipping, falling, etc.), hand gestures (e.g., wave, push, pull, sweep, clap, slide, draw the circle, draw zigzag, sign languages, digit finger gestures, etc.), and exercises (e.g., table tennis action, bodyweight exercise), etc. Differently, activity inference usually refers to determining the meaning of actions. This application aims to make the behavior inference to reveal the real purpose of the activities instead of specific behaviors, such as smoking activity detection, crowd counting, and user authentication, etc.

In general, according to identification techniques, CSI-based recognition applications can be divided into two types: pattern-based applications and model-based applications. The former recognizes human behavior by leveraging pattern recognition methods which usually include some machine learning algorithms. In addition, the pattern-based approach usually requires more data to fit or train network parameters. Differently, the latter recognizes behavior by utilizing the mathematical or physical model that depicts the unique relationship between CSI signal variation and human behaviors. Thereby, the key of model-based methods

is how to extract effective signal features to express human behavior and build a model to measure the signal changes caused by human activity. From the description of the model-based approach, this approach determines human activities with the model without the need of a lot of data, which is its striking advantages compared with the pattern-based method. Recently, deep learning technology has been widely applied in various fields due to special characteristics compared with traditional pattern-based methods. Specifically, it requires a great number of experiment data to train neural network parameters because it usually contains tens of thousands to several million unknown parameters. Besides, the deep learning-based approach has no definite feature extraction procedure because this step usually is integrated into different network structures. In other words, the deep learning-based method can automatically extract or build complex features by middle lays, which is an apparent difference compared with pattern-based methods. Although the deep learning-based methods usually are considered as one of the pattern-based, based on the above analysis, we deem that deep learning applications should be classified as a group because of its distinct difference compared with pattern-based applications. Consequently, this paper categorizes state-of-the-art applications into three groups, including model-based, pattern-based, and deep learning-based approach.

A. PATTERN-BASED BEHAVIOR RECOGNITION

In CSI-based behavior recognition applications, we find more applications based on the pattern. The pattern-based approach aims to find distinct patterns to identify human behavior. This depends on effective feature construction that can be leveraged to recognize different activities [3]. These features can be either complicated or simple on the basis of behavior granularity and recognition purposes. Thereby, the pattern-based approach usually need some samples to construct patterns to accomplish CSI-based behavior recognition applications. After collecting effective samples, we need to preprocess the CSI data to remove the error data and noise data from hard devices and ambient scenarios. The typical signal preprocessing techniques include low past filter, Hamel filter, Principal Component Analysis (PCA), Discrete Wavelet Transformation (DWT), and Butterworth filter, etc. Especially, for phase information, we must implement phase calibration because it has a good resolution for human action and is distorted easily by device imperfection and other disturbance. Afterward, we obtain clear data and can feed them into a classifier to implement behavior recognition.

Currently, the general classifiers include Dynamic Time Warping (DTW), Support Vector Machine (SVM), k-Nearest Neighbor (KNN), Hidden Markov Model (HMM), BP Neural Network (BPNN), Self-Organizing Map (SOM), etc. This article presents some typical behavior recognition applications, including daily behavior recognition [55]–[62], table tennis action recognition [63], bodyweight exercise recognition [64], danger pose detection [65], abnormal activity detection [66], falling detection [67]–[71], hand gesture recognition [72]–[81], sign language recognition [82], [83], sleep monitoring [84], respiration detection [85]–[88], lip reading and speech recognition [89], [90], keystroke detection [91], [92], writing recognition [93]–[95], sedentary behavior monitoring [96], smoking detection [97], crowd counting [98]–[103], step counting [104], [105], human presence detection [106]–[111], and user authentication [112]–[123].

B. MODEL-BASED BEHAVIOR RECOGNITION

In the typical CSI-based behavior recognition applications, we find that many different behavior models have been employed. The model-based approach aims to design a physical model that relates the signal space to the physical space and builds the physical law through the relationship between the received signal and the sensing target [3]. In other words, the model-based recognition approach correlates the user's motion to the received signal variation. Therefore, it can recognize human behavior by exploring the physical law based on the physical-mathematical models. Instead of profile matching or pattern recognition, we can utilize the model to identify human behavior with low computation cost. Compared with the pattern-based method, the model-based approach can achieve satisfactory performance with small measurement data because of its utilization of the model. Despite these advantages, it is challenging to apply the model-based approaches in some scenarios because test environments usually contain various stuff, making it hard to develop a general mathematical model that depicts the CSI features in the scenario.

Although many references involve model-based methods, they do not give an explicit description. In this article, we consider two types of models used in CSI-based behavior recognition applications. The first one exploits the physical law of signal propagation to depict the relationship between the signal change of CSI and the user's behavior, which can be used to recognize human activities, such as Fresnel Zone Model, angle of arrival (AoA), CSI-ratio model, interacting model. Therefore, we can obtain the human action state from these models, such as motion direction, speed, and time, which interprets human behavior by using physical law solely. The second one usually comprises two parts, the physical description of signal changes and recognition algorithms, such as the CSI-speed model and CSI-activity model. Specifically, this method first obtains signal fluctuations based on propagation law and physical model and then leverages classification or fitting methods to identify human behavior. Therefore, this method combines models and classification algorithms. Remarkably, we do not consider PSD (power spectral density) and CFR (channel frequency response) as a model because they solely describe CSI signal changes based on the propagation path and cannot be utilized to depict behaviors directly. This paper presents some representative model-based recognition applications, including daily behavior recognition [124], [125], human presence detection [126], [127], user authentication [128], walking direction

estimation [129], handwriting recognition [130], and respiration detection [24], [131]–[137].

C. DEEP LEARNING-BASED BEHAVIOR RECOGNITION

Recently, deep learning technology has been studied comprehensively and has achieved considerable progress. Therefore, many recognition applications based on CSI using deep learning algorithms have been developed. Although the number of traditional pattern-based approaches of CSI exceeds that of deep learning-based approaches, the growth speed of deep learning applications is rapidly increasing. Compared with pattern-based and model-based approach, deep learning-based approach has many distinct benefits. We can benefit from automatic feature discovery and behavior recognition of deep learning algorithms. Specifically, it can automatically identify and extract effective features and feed them into the classifier. The procedures of feature extraction and classification are implemented by the middle layers of the deep neural network model. To accurately depict the features, the deep learning-based approaches require a large number of samples to continuously alter network parameters. The procedure of training is labor-intensive and time-consuming, which is its main disadvantage.

For CSI-based behavior recognition, we can obtain satisfied identification accuracy if we build an appropriate model and have massive samples. That is to say, there is no need to pre-process CSI data that has complex features to obtain feature descriptions in deep neural networks, which is the advantage of the deep learning-based method. Besides, we have to collect plenty of samples of CSI to train network parameters and obtain good recognition accuracy. This article presents some typical deep neural network models used in CSI-based behavior recognition applications, including Autoencoder, Convolutional Neural Network (CNN), LSTM, RNN, Residual Neural Network (ResNet), and Restricted Boltzmann Machine (RBM). In addition, we also detailed introduce these applications, such as daily behavior recognition [138]–[142], falling detection [143], [144], syncope detection [145], hand gesture recognition [8], [146]–[149], sign language recognition [150], gait and walking direction recognition [151], [152], human presence detection [153]–[155], crowd counting [156]–[159], user authentication [160]–[163], and respiration monitoring [22].

The main contributions of this paper are given as follows. Firstly, we review the latest research progress of WiFi CSI-based human behavior recognition and emphasize the size of the samples. To the best of our knowledge, the present paper is the first survey that contains a total sample description of CSI-based behavior identification. We exhibit the accurate size of samples, classifiers and human behavior from pattern-based, model-based, and deep learning-based approaches according to identification techniques. Secondly, we present the general framework that contains the main structure of a typical recognition system. Based on the structure, we report a comprehensive statistic of state-of-the-art applications from base signal selection, signal preprocessing and behavior classification methods. Thirdly, we investigate the state-of-the-art studies and exhibit them from the different granularity of behavior. We select some typical applications and outline their characteristics from test equipment, preprocessing methods, test environments, the size of participants, recognized behaviors, classifiers, and system performance. Finally, we make a comprehensive discussion about a few typical activities from the implementation view and provide some helpful insight into the development of these applications.

The remainder of this article is arranged as follows. In Section II, we present some surveys of human behavior recognition. Moreover, we introduce background which contains the basic concept of CSI and the principle of CSI-based behavior recognition. In Section III, we give some general approaches and the general framework of CSI-based behavior recognition, including base signal selection, signal preprocessing, and three behavior recognition techniques (e.g., pattern-based recognition, model-based recognition, and deep learning-based recognition). In Section IV, we divide state-of-the-art CSI-based activity identification applications into three groups and many subgroups based on behavior and review them in detail. Section V presents a discussion about identification techniques and system performance. Issues and future directions are given in Section VI. Finally, the conclusion is provided in Section VII.

II. RELATED WORK, CHANNEL STATE INFORMATION, AND THE PRINCIPLE OF CSI-BASED BEHAVIOR RECOGNITION

A. RELATED WORK

With the wide popularity of wireless networks, WiFi devices have been extensively employed in various indoor environments. Currently, WiFi-based behavior recognition approaches have been widely studied and the related applications are increasingly emerging due to their major strengths. Meanwhile, some surveys of WiFi-based recognition report the behavior recognition progress using CSI and review the state-of-the-art research results, (see in Table 1). However, they have some weaknesses in some aspects. Next, we summarize their major features. Although reference [164], [169] present the survey on WiFi-based contactless activity recognition and highlight behavior recognition frameworks, they do not provide enough description of the typical applications. Liu *et al.* [173] review wireless sensing applications using 4 different techniques, including RSSI, CSI, FMCW (Frequency Modulated Carrier Wave), and Doppler shift. The authors select some typical applications and analyze the features of them. However, this paper lacks the description of signal preprocessing, experimental performance evaluation and recognition approach analysis. The articles [3], [165]–[167] come from the special issues of behavior recognition based on Wi-Fi CSI and present the survey on CSI-based behavior recognition from a different view, including pattern-based and model-based approaches, deep learning

TABLE 1. Related surveys on CSI-based behavior recognition.

Paper	Main Work	References	Samples (train/test)	Year
Ma *et al.* [164]	recognition framework	50	no mention	2016
Wu *et al.* [3]	pattern-based, model-based	15	mention, insufficiency	2017
Yousef *et al.* [165]	deep learning techniques	15	mention, insufficiency	2017
Wang *et al.* [166]	behavior granularity	15	mention, insufficiency	2018
Jiang *et al.* [167]	principles, capabilities and limitations	69	no mention	2018
Wang *et al.* [168]	existing models and features	15	no mention	2018
Lu *et al.* [169]	recognition framework	85	no mention	2018
Khalili *et al.* [170]	different applications	167	mention, insufficiency	2019
Ma *et al.* [171]	more applications	157	mention, insufficiency	2019
Wang *et al.* [172]	applications of through-the-wall scenario	97	mention, insufficiency	2019
Liu *et al.* [173]	FMCW, CSI, RSS, Doppler shift	138	no mention	2019

techniques, and behavior granularity. Although these articles present the key content from their interest, they lack the detailed analysis of typical behavior recognition applications. Wang *et al.* [168] present a survey on WiFi-based recognition from challenges, opportunities, and applications, and emphasize the model used in WiFi-based recognition. Although it describes some physical models and discusses further research opportunities, it does not provide a comprehensive analysis of existing applications. Two papers [170], [171] are the newest surveys that cite more references and present a detailed survey on WiFi-based recognition. However, Khalili *et al.* [170] do not introduce the framework of activity recognition and reference [171] lacks a detailed description of every application. Although reference [172] investigates many references, it focuses on through-the-wall applications and emphasizes the special application scenario, whose aim is different from that of this paper.

Based on the above analysis, we find that these surveys do not present a comprehensive review of WiFi CSI-based sensing, especially for human behavior recognition. Therefore, there is a pressing need for a detailed survey on CSI-based behavior recognition that contains general activity recognition framework, more application overview, more detailed signal processing statistics, accurate classification approach description, and the statement of typical applications in detail.

This article effectively overcomes the weaknesses existed in the above surveys and contains some essential contents, including the accurate number of samples, detailed preprocessing method statistics, more comprehensive application analysis, and more thorough references. Specifically,

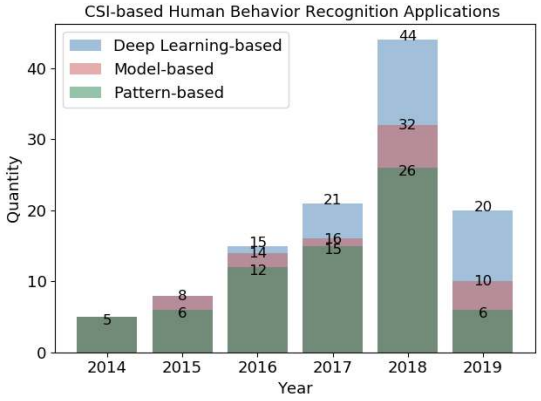

FIGURE 1. Recognition technique statistics of the different time from Table 3.

this paper investigates more than 100 latest behavior recognition applications using CSI and categorizes them from behavior recognition methods. Besides, we make statistics about the number of different recognition methods according to the paper publishing time, as shown in Fig. 1. The number of behavior recognition applications based on CSI goes up rapidly, which indicates that this is a heat research topic. We notice that the numbers of pattern-based and deep learning-based behavior recognition applications are gradually increasing. The growth rate of deep learning-based applications is quicker than that of the pattern-based applications. The number of model-based applications quickly increases in 2018 largely because the Fresnel zone model is proposed and applied, which shows that this approach can effectively monitor the breathing rate. From Fig. 1, we deem that the number of behavior recognition applications based on CSI will steadily rise because of the wide adoption of pattern-based methods, further study of deep learning-based method, and proposal of the appropriate model.

To clarify the major content of this paper, we use some tables to summarize the analysis results of human recognition using CSI. Specifically, the paper exhibits the behavior recognition framework using CSI and main composition (see in Table 2), including base signal selection, signal preprocessing, and three types of classification methods. This paper presents the state-of-the-art studies (see in Table 3) and analyzes these CSI-based behavior recognition applications from classification technique aspects, such as the pattern-based, the model-based, and the deep learning-based. For a more detailed analysis of these applications, this article presents Tables 4 to 8, which contains system name, equipment, preprocessing methods, experimental environments, the number of participants, classifiers, behaviors, and system performance, etc. In addition, this article presents a discussion which recognition approach (e.g., the pattern-based, model-based, and deep learning-based) is more suitable and has the best performance for some typical behaviors, such as daily behavior, falling detection, hand gesture, crowd counting, user authentication, respiration monitoring. In sum, this paper presents a complete framework, precise sample

statistic, detailed analysis of applications, and a comprehensive discussion about CSI-based behavior recognition applications.

B. CHANNEL STATE INFORMATION

CSI is a metric that describes the channel properties of wireless communication links and considers the several factors affecting signal propagation, such as signal scatter, environmental attenuation, and distance attenuation. The purpose of the introduction of CSI is to ensure effective and reliable data transmission by quantifying the channel fading effect and adjusting the signal transmission rate. Specifically, when the wireless signal propagates in a multi-path manner, it will be obstructed by the objects in the line of sight (LOS) path, which leads to signal changes, including amplitude attention and phase shift. Besides, the reflection from the test environment also changes the signal waveform. Therefore, CSI is introduced to evaluate the communication link state. That's to say, the quality of the wireless channel can be estimated by the CSI matrix, and the communication rate can be adjusted based on the CSI. In the IEEE 802.11n/ac standards, CSI is measured and parsed from the PHY layer using orthogonal frequency division multiplexing (OFDM) technology. In the frequency domain, the wireless channel can be defined as:

$$\mathrm{Y} = \mathrm{H} \times \mathrm{X} + \mathrm{N} \tag{1}$$

where H is the channel matrix representing CSI information; the received and transmitted signal vectors are Y and X, respectively; N refers to an additive white Gaussian noise vector. According to formula (1), H can be expressed as:

$$H\,(i) = |H(i)|\, e^{j sin \angle H(i)} \tag{2}$$

where $H(i)$ represents the value of CSI for the i^{th} subcarrier, which includes the amplitude and phase of CSI; the amplitude and phase of the i^{th} subcarrier are $|H(i)|$ and $\angle H\,(i)$.

C. THE PRINCIPLE OF CSI-BASED HUMAN BEHAVIOR RECOGNITION

The principle of WiFi CSI-based behavior recognition is based on the observation that different human motion within the coverage of the WiFi signal can cause distinct wireless channel disturbances and the corresponding relationship is unique. Therefore, we can utilize the correlation between CSI dynamics and user activities to recognize human behaviors. To effectively describe wireless signal changes caused by human behaviors, we explain how the radio signal propagates and why the users affect wireless channels as follows.

A typical indoor human behavior recognition system consists of a WiFi access point, PC (e.g., WiFi receiver), and users. The wireless signal propagates in a multi-path manner and the wireless channel keeps relatively stable when there are no people in the range of the WiFi signal. However, once a person stands or moves in the coverage, the signal propagation paths will change due to signal reflection, which results in channel disturbances, including amplitude attention and phase distortion. In other words, the users will cause the received channel state information changes due to the multi-path effects, which can be leveraged to recognize human behaviors. Specifically, the unique corresponding relationship between CSI dynamics and human behaviors can be obtained based on the behavior profile construction, which can be utilized to recognize different behaviors by correlating the signal changes with the corresponding channel distortion patterns. As shown in Fig. 2, according to the Friis [174] free space propagation equation, the power receiver can be represented as follows.

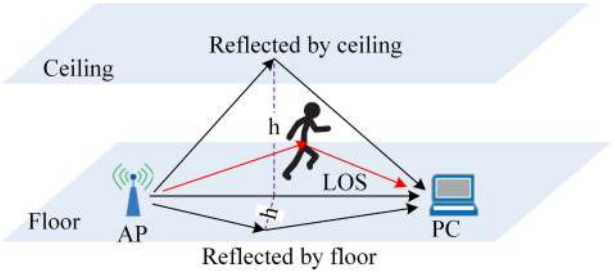

FIGURE 2. WiFi signal propagation.

$$P_r = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 (d + 4h + \Delta)^2} \tag{3}$$

where λ is signal wavelength; d is the length of the LOS between AP and PC; P_t and P_r refer to the transmitting and receiving power, respectively; G_t and G_r are the transmitting and the receiving gains, respectively; h is the vertical distance from the reflection point to the LOS; Δ is the length of the body's reflection path; h and Δ are crucial factors that affect the received power, which indicates that the obstacle and reflection significantly affect signal propagation. From the formula (3), the received power is inversely proportional to the square of the propagation path. That is to say, the greater the distance, the smaller the receiving power and the more difficult the sensing.

In the frequency domain, the channel frequency response (CFR)$H(f, t)$ can be defined as follows.

$$Y\,(f, t) = H(f, t) \times X(f, t) \tag{4}$$

$$H\,(f, t) = e^{-j2\pi \Delta f t} \sum\nolimits_{k=-1}^{N} a_k\,(f, t)\, e^{-j2\pi \tau_k(t)} \tag{5}$$

where $Y\,(f, t)$ and $X(f, t)$ are the received and transmitted signal, respectively; $a_k(f, t)$ refers to the signal changes of the k^{th} path; the phase shift caused by the propagation delay of $\tau_k(t)$ is $e^{-j2\pi \tau_k(t)}$; $e^{-j2\pi \Delta f t}$ refers to the phase shift caused by frequency offset Δf between the transmitter and the receiver.

As shown in Fig. 3, the human's movement changes the WiFi signal propagation; thus, the propagation paths can be divided into two categories, such as static path and dynamic path. In general, the signal of the static path is not affected by the human body and the frequency state response of this path is $H_s(f, t)$. Differently, the signal of the dynamic path is affected by the human body and other objects (e.g., floor,

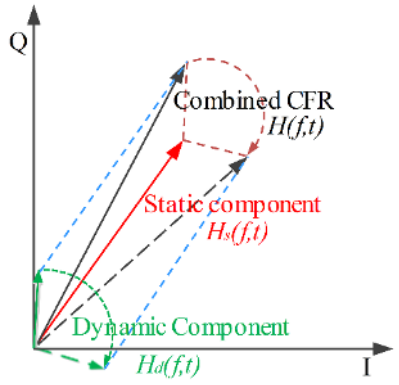

FIGURE 3. Multi-paths caused by behaviors.

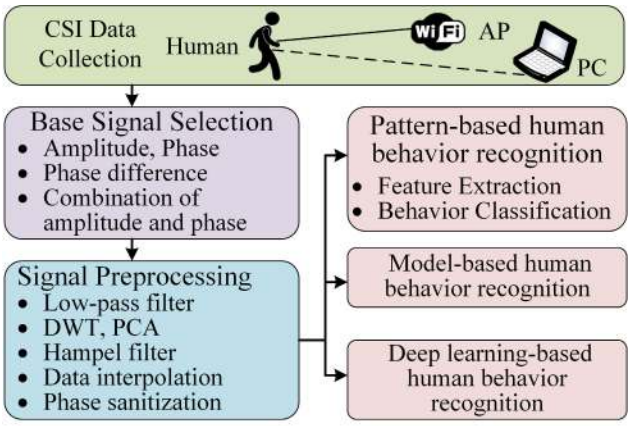

FIGURE 4. CSI-based behavior recognition framework.

wall, etc.), and the dynamic path frequency state response $H_d(f,t)$ can be expressed as follows.

$$H_d(f,t) = \sum_{k\in P_d} a_k(f,t)exp\left(-j2\pi d_k(t)/\lambda\right) \quad (6)$$

where P_d refers to a total dynamic path; $d_k(t)$ is the length of the k^{th} path at time t; λ is the wavelength. The total CFR power is expressed as follows.

$$H(f,t) = e^{-j2\pi\Delta ft}(H_s(f,t) + H_d(f,t)) \quad (7)$$

The energy of the entire CFR changes along with the body's motion; therefore, we can make the following assumption that the human body moves with a constant speed, the length of a certain path k changes at a constant speed v_k for a short time period. Let $d_k(t)$ denotes the length of the k^{th} path and $d_k(t) = d_k(0) + v_k t$, the instantaneous CFR power in time t can be written as:

$$\begin{aligned}|H(f,t)|^2 &= \sum_{k\in P_d}|a_k(f,t)|^2 + |H_s(f,t)|^2 \\ &+ \sum_{k,l\in P_d,k\neq l} 2|a_k(f,t)a_l(f,t)| \\ &\times\cos\left(\frac{2\pi(v_k - v_l)t}{\lambda} + \frac{2\pi(d_k(0) - d_l(0))}{\lambda} + \phi_{kl}\right) \\ &+ \sum_{k\in P_d} 2a_k(f,t)H_s(f,t) \\ &\times\cos\left(\frac{2\pi v_k t}{\lambda} + \frac{2\pi d_k(0)}{\lambda} + \phi_{sk}\right) \end{aligned} \quad (8)$$

where the initial phase deflection of the signal is $\frac{2\pi d_k(0)}{\lambda} + \phi_{sk}$ or $\frac{2\pi(d_k(0)-d_l(0))}{\lambda} + \phi_{kl}$. Formula (7) and formula (8) provide a significant view: the total CFR power is a sum that contains a constant offset and a set of sinusoids, where the frequencies of the sinusoid are related to the speed changing rate of path length. By measuring these sinusoid frequencies and multiplying the carrier wavelengths with these frequencies, the speed variation of the path length can be obtained. Then the movement speed of the human body can be calculated by observing the CSI changing, and different human behaviors can be identified.

III. GENERAL METHODS AND FRAMEWORK OF CSI-BASED HUMAN BEHAVIOR RECOGNITION

In this section, we present a general framework for CSI-based behavior recognition using COTS (commercial off-the-shelf) WiFi devices, introduce the crucial components, interpret their important role, and analyze their key characteristics. Based on signal processing procedures, the behavior recognition system comprises four parts, such as data collection, base signal selection, signal preprocessing, and behavior classifiers, as shown in Fig. 4. Based on the behavior recognition technique, we first categorize the recognition approaches into three groups, including pattern-based, model-based, and deep learning-based methods. Then, we introduce the key parts of the classifier of three types and analyze their main characteristics.

In general, we collect CSI data using a PC equipped with a network interface card (NIC) when the PC communicates with the access point (AP). After collecting CSI data, we must choose a suitable base signal, such as amplitude, phase, the combination of amplitude and phase, or phase difference. In order to wipe out the noise of the CSI stream and obtain more accurate CSI data, signal processing approaches become essential. As shown in Table 2, this paper presents some representative techniques to acquire precise CSI data, including low-pass filter, Hampel filter, PCA, DWT, data interpolation, phase sanitization, Butterworth filter, Savitzky-Golay filter, and Band-pass filter, etc. After signal processing, we recognize human behaviors using three types of recognition techniques, such as pattern-based, model-based, and deep learning-based approaches. The pattern-based method usually needs to extract features and classify different activities using machine learning algorithms (e.g., DTW, SVM, KNN, HMM, etc.) and neural networks (e.g., BPNN, SOM). The model-based method usually utilizes some typical models, such as Fresnel zone model, AoA, human respiration model, interacting model, CSI-speed model, and CSI-activity model, to identify human behaviors. Compared with the pattern-based methods, the model-based approach has fewer applications due to the difficulty of the building model.

TABLE 2. CSI-based human behavior recognition process.

Process		Technique
Base Signal Selection		Amplitude [55], [61], [63], [65], [67], [70], [72], [74], [76], [83], [93], [108], [112], [113], [115], [118]-[122], [125], [126], [132], [134], [138], [141], [144], [156], [160], [161], Phase [57], [59], [73], [78], [80], [87], [97]-[99], [106], [108], [124], [136], [146], [175], Combination of amplitude and phase [24], [56], [66], [67], [75], [79], [88], [96], [112], [129], [131], [139], [147], [150], [157], [176], Phase difference [8], [68], [69], [100], [102], [126], [136], [143], [85], [86].
Signal Preprocessing		Hampel filter [4], [84], [117], [134], [143], [177], PCA [57], [59], [66], [91]-[93], [96], [105], [114]-[116], [124], [138], [157], [178], DWT [57], [76], [93], [105], [115], Data interpolation [68], [97], [108], [143], Low-pass filter [55], [68], [69], [74], [82], [83], [91], [108], [138], Phase Sanitization [110], [157], [165], Phase calibration [65], [146], [86], MCS index filter [55], LOF [67], [73], SSA [143], Butterworth low-pass filter [63], [74], [115], [161], Birge-Massart filter [73], median filter [82], [132], [179], [180], spikes removal filter [83], wavelet filter [84], [99], [132], Butterworth IIR band-pass filter [89], Butterworth filter [91], [93], [112], [113], [157], FFT [96], [100], [135], PEM [98], THD [100], CWT [116], STFT[114], [161], TOA filter [132], Savitzky-Golay filter [125], [129], adaptive filter [22], Band-pass filter [146].
Pattern-based	Feature Extraction	Statistical characteristics [57], [59], [61], [68], [69], [73], [93], [113], [116], [158], [181], Doppler Shift [48], [64], [75], [111], [114], [182], [183], Wavelet features [108], [124], [184], Time-frequency diagram [55], [56], [63], [76], [96], [97], [105], [112], [124], [138], [179].
	Behavior Classification	DTW [55], [74]-[76], [89], [93], [113], [115], [178], SVM [57], [68], [69], [73], [83], [99], [107], [108], [114], [120], [146], KNN [58], [76], [91], [93], [115], HMM [42], [109], [124], BPNN [16], [61], SOM [60], [94], Bayesian model [66], [111], Random Forest [67], Semi-supervised learning [83], [100], transfer learning [83], SVD [84], PPMCC [96], SRDA [89], MCFS [89], SVD [84], PPMCC [96], K-means clustering [4], SRDA [89], MCFS [89], MUSIC [107], [126], SRC [56], [185].
Model-based		Fresnel zone model [24], [125], [126], [129], [131]-[135], AoA [107], [126], [130], [136], [176], Human respiration model [131], Interacting model [128],CSI-ratio model [137], Gaussian Mixture Model [127],CSI-speed model [124], [186], CSI-activity model [124].
Deep Learning-based		Autoencoder [158], [160], DCNN [143], [152], CNN [8], [22], [138], [141], [147], [150], [161], [187], [188], LSTM [138], [141], [165], [178], RNN [141], [151], [178], BiRNN [151], AE-LRCN [139], [140], ResNet [145], [161], RBM [146].

At present, the deep learning-based method is gradually applied in activity identification. Although this approach requires a large number of samples to train a deep neural network (e.g., Autoencoder, CNN, LSTM, RBM, etc.), it not only can automatically extract behavior features and classify samples but also can achieve satisfactory performance. In summary, this architecture gives us a general description of CSI-based behavior recognition. The more detailed activity recognition processes can be illustrated as follows.

A. BASE SIGNAL SELECTION

Generally speaking, accurate and complete data collection depicting user movements is very important to effective behavior recognition. Without adequate data, it is hard to mine behavior profiles or build a mathematical model to characterize the correlation between human behavior and corresponding signals. Furthermore, the selection of the base signal plays a crucial role in behavior recognition as it decides the resolution of data and the accuracy of behavior identification. For example, a complex number contains more information compared to a real number as we may calculate the direction and amplitude based on its real and imaginary parts. Currently, we can categorize the signal selection of existing applications into four groups: amplitude, phase, the combination of amplitude and phase, and phase difference.

1) AMPLITUDE

The physical meaning of CSI amplitude is the quantification of the signal power attenuation after multi-path fading [186]. The user motion in the WiFi-enabled area affects the wireless signal propagation and changes the amplitude of signal arriving at the receiver, leading to amplitude variation. There is a unique relationship between the frequency of amplitude changes and human walking speed, which means the movement of a person can be detected after the measurement of amplitude. For instance, E-Eyes [55] utilizes amplitude measurements of CSI to realize nine kinds of daily activity (e.g., walking, cooking, washing, etc.) identification. FallDeFi [70] achieves 93% precision of falling detection through the analysis of amplitude variation of CSI subcarriers. WiFi-ID [112] realizes user identity recognition with 93% accuracy for two users by analyzing the amplitude of the CSI stream.

2) PHASE

In theory, the phase contains more information than the amplitude and can be used to depict the signal changes and corresponding user motion. However, as phase is periodic compared to amplitude and its measurement value is affected by device clock and carrier frequency, it must be calibrated to generate practical phase features. Usually, a linear transformation is a simple and effective calibration method. For example, Wu *et al.* [106] employ the linear transformation to revise phase information of CSI and implement user authentication. However, phase calibration has some adverse effects. For instance, a key limitation of the phase calibration is the lack of clear explanation and accurate description, which leads to the difficulty of building a precise phase-based model that represents the user motion. Meanwhile, the filter process might remove some phase information describing user movement, resulting in the loss of some detection capability.

3) COMBINATION OF AMPLITUDE AND PHASE

A combination of amplitude and phase might be a useful approach in that it can effectively utilize the advantages of both features and improve recognition accuracy. For instance, WiseFi [176] achieves better performance (e.g., average accuracy: LOS 89.1%, NLOS 82.5%, and through-one-wall 73.4%) for activity recognition by using the combined approach. NotiFi [66] can automatically detect abnormal activities by applying the combination of phase and amplitude of CSI. SignFi [150] identifies nine digits finger gestures with an average 86.66% precision by utilizing the information of phase and amplitude.

4) PHASE DIFFERENCE

Compared to phase, phase difference has several benefits such as space diverse and less noise. It is sensitive to human movements and can be leveraged to identify behaviors. Currently, there have been several studies that use the phase difference to detect various vital signals and complex activities. For example, Anti-Fall [68] distinguishes fall and fall-like activities by using the CSI phase difference over two antennas. RT-Fall [69] applies the phase difference of CSI to realize falling detection with satisfactory performance. FreeSense [115] can automatically detect humans by the analysis of the phase difference of two signals. PhaseBeat [85] can monitor the user's respiration rate and heartbeat by analyzing the CSI phase difference. However, the approach usually uses an antenna array to measure phase difference, which leads to this method unavailable without an antenna array.

B. SIGNAL PREPROCESSING

As known to all, the measurements of CSI comprise helpful signals, disordered noise, and a few outliers. These bad data mainly come from ambient environments, signal interference, and some moving persons. Thus, the methods of data preprocessing are significant since it can remove outliers, filter noise, calibrate the phase, and retain valuable data. Currently, various signal preprocessing techniques can be used to implement noise removal, data interpolation, and phase sanitization. Generally speaking, the premise of correct behavior recognition is to obtain precise data representing human behaviors. Therefore, the data preprocessing is necessary before further feature extraction. Specifically, to reduce invalid data and enhance the accuracy of behavior recognition, we must remove the noise caused by multi-path effects and hardware devices. There are several de-noising methods, such as low-pass filter, Hamel filter, PCA, Discrete Wavelet Transform (DWT), data interpolation, and phase sanitization.

1) LOW-PASS FILTER

The low-pass filter is an ordinary method that only allows some signals below the cutoff frequency to pass. Since human activities always cause low variation frequency of CSI signals, the low-pass filter can effectively wipe out high-frequency noise and constant value that cause by the multiple-path effect and has been widely used at many applications, such as RT-Fall [69], Anti-fall [68], WiGeR [74], and R-TTWD [108], etc. Although the low-pass filter has great performance, there are still some noises that cannot be removed effectively. These noises usually come from the internal state of WiFi network interface cards (NICs) of sender and receiver or environment variation [169].

2) HAMPEL FILTER

Currently, there have many abnormal values in the collected CSI data due to the sudden changes in equipment and environment, and so on. Hampel filter is an efficient technique to remove outliers that are far away from their neighbor data. Specifically, it finds the outliers and replaces them with the mean of data by utilizing a moving average window, which eliminates the negative impacts caused by invalid data. In the case of the CSI signal, the outliers caused by the intrinsic hardware feature and deployment environment can be wipe out by Hamel filter. For instance, Wi-Sleep [84] applies Hamel filter to wipe out the noise and monitors respiration rates. WiHACS [57] utilizes Hamel filter to denoise and identifies seven kinds of activities, i.e., sitting, walking, falling, etc. Although Hamel filter is efficient to remove noise, this approach usually fits these signals that only have Gaussian noise [164].

3) PRINCIPAL COMPONENT ANALYSIS

Principal Component Analysis (PCA) is a data processing method that represents the data with part features of a matrix, which reduces data dimensionality and enhances processing efficiency. Specifically, some eigenvectors are selected to construct a new matrix that represents the original matrix by computing the eigenvalue of the matrix. In CSI-based behavior recognition, the PCA is mainly used to wipe out the noise and data redundancy at signal processing steps (e.g., Wi-Wri [93], WiHACS [57], FreeSense [126], and CareFi [96], etc.). The PCA works well if the selective principal components can describe the original matrix [189]. In the case of CSI data processing, we can employ PCA on the measurement data to remove various noises and redundant data coming from off-the-shelf devices and environments.

4) DISCRETE WAVELET TRANSFORM

Discrete Wavelet Transform (DWT) can be used at image processing because it can remove the noise of signal, extract and preserve some useful image edge information. It can overcome the shortcomings of the traditional Fourier transform image processing. These shortcomings include difficulties in detecting the local abrupt signals and the loss of image edge information when removing signal noise, and so on. On the contrary, DWT performs better than the Fourier transform in removing signal noise. Currently, there has a lot of applications that utilize DWT to denoise, such as WiFinger [76], WiStep [105], etc.

5) DATA INTERPOLATION

During the realistic applications, despite a constant rate of data delivery at the transmitter, the receiver usually cannot receive stable and constant data due to data loss or transmission delay. Therefore, effective means must be taken to address this question. Data interpolation is an effective method to solve this question due to its simplicity and practicality. Specifically, to get data at a fixed rate, we can add the sample computed from its neighbor to the time slot missing data, which constructs even CSI sequences, eliminates data clustering and fuzzy measurements, and enhances data accuracy. For example, Smokey [97] employs the interpolation method to preprocess the received data and achieves the recognition of smoking behavior.

6) PHASE SANITIZATION

Although we can acquire the raw phase information, it cannot be used directly due to carrier frequency offset (CFO) and sampling frequency offset (SFO) [165]. Specifically, CFO comes from the lack of synchronization at the central frequency between the transmitter and receiver clocks, and the process of the analog-to-digital converter (ADC) can be likely to produce SFO. Moreover, the raw phase P_M can be expressed as [157]:

$$P_M = P + 2\pi \frac{m_i}{N} \Delta t + \beta + N \tag{9}$$

where P is the genuine phase; Δt is the time lag due to SFO; β is the phase offset due to CFO; N is the noise. From formula (9), due to the value of Δt and β are unknown, the genuine phase cannot be calculated directly. However, phase sanitization (e.g., linear fitting method) is proposed to remove the impacts of CFO and SFO, which makes phase information applicable for any detection [110].

C. PATTERN-BASED HUMAN BEHAVIOR RECOGNITION

The purpose of pattern-based recognition is to find the patterns and regularities from data. In most cases, the recognition procedure is to assign a label to a given input data. Therefore, many recognition problems can be treated as classification problems. As for pattern-based behavior recognition, we attempt to identify human behaviors by leveraging CSI variation pattern. Specifically, we first collect data that contains CSI changes caused by human behaviors. Then, we determine the CSI variation regularity and establish a unique mapping relationship between CSI variation and human behaviors. Based on the mapping, we exploit some recognition methods to identify specific behaviors. Therefore, the key to this approach is to relate the CSI signal variation profile to specific actions. How to identify and describe distinct patterns to distinguish different behaviors is challenging because CSI change is complicated. In general, the pattern-based approach for behavior recognition consists of two processes: feature extraction and activity classification. From identification processes, we can discover that these two steps are crucial to accurately identifying human behaviors. We introduce their basic principles and main implementation methods as follows.

1) FEATURE EXTRACTION

Feature extraction refers to a process that usually transforms the original information into new data types that can be easily utilized by the following classification. After feature extraction, we get a more effective data description. The measurement data usually contain much redundant information and they cannot be utilized directly without feature extraction. Therefore, feature extraction plays a crucial role in CSI recognition approaches. Currently, we can extract many features from preprocessed data, such as statistical characteristics, Doppler shift, wavelet features, and time-frequency diagram.

a: STATISTICAL CHARACTERISTICS

Statistical characteristics refer to some statistical results computed from the original waveform data. They effectively describe the general features, simplify data representation and reduce the data complexity. We can calculate the statistical features based on the original time-domain data. Similarly, we can also get the statistical results based on the frequency domain data after FFT transformation of time-domain data. Some behavior recognition applications employ statistical features to extract features, such as TW-See [61], Anti-Fall [68], RT-Fall [69], etc.

b: DOPPLER SHIFT

Doppler shift refers to the frequency changes of the original signal observed at one user when there is a relative speed between transmitter and user. It usually can be used to extract frequency shifts and deduce the object speed. Specifically, human movement changes the path length of the body's reflected signal, resulting in changes in the signal at the receiving end. The Doppler frequency shift of the received signal can be expressed as follows.

$$f_{Doppler} = f \frac{v_{path}}{C} \tag{10}$$

where c is the speed of light propagation; v_{path} refers to the speed of the path length changes; f is the carrier frequency of the signal. Currently, there have some activity recognition applications based on Doppler shift. For instance, WiSee [182] implements behavior recognition by exploring the Doppler shift. WiFit [64] counts bodyweight exercise repetitions by applying the doppler shift.

c: WAVELET FEATURES

Wavelet transform is a new feature transformation method similar to the short-time Fourier transform (STFT). It overcomes the shortcoming of STFT that window size does not change along with frequency differences. After wavelet transform, the wavelet coefficients of each frequency band are the corresponding features needed to be processed. We call these coefficients as wavelet features. Wavelet transform achieves better performance of local feature extraction as it can

analyze signals at the scale of more fine-grained frequency. For example, the CARM [124] uses the wavelet transform to distinguish speeds and frequencies of different behaviors and realizes eight kinds of behaviors (e.g., running, walking, sitting, and falling) recognition with an average accuracy of 96%.

d: TIME-FREQUENCY DIAGRAM

The time-frequency diagram describes the relationship between time and signal amplitude of all frequencies. For a given time signal, we can transform it into a two-dimensional color image using a time window. The image contains the time, different frequencies, and corresponding amplitudes. In the case of CSI signal, the image illustrates the time index, subcarrier index, and signal amplitude values reflected by human motions. Based on the energy changes of different frequencies along with time, we can clip the time, get each action segment and identify specific behaviors, such as E-eyes [55], WiTT [63], Smokey [97], and WiStep [105], etc.

2) BEHAVIOR CLASSIFICATION

After feature extraction, we will identify human behavior by utilizing different classifiers based on the characteristics of specific applications. Specifically, for some simple classification problems that can be represented by a statistical characteristic, we can use the simplest comparison method by designing the threshold to distinguish the behavior types. In other words, we can acquire a computing value based on some methods, and it can be used to determine whether the behavior is identified to a certain type according to the threshold assigned. For example, Kun *et al.* [190] use this method to recognize whether there are human activities in the environment and Zhang *et al.* [68] compare the standard deviation of signal with the threshold to determine whether a ''fall-like activity'' is a fall action.

Differently, for the complex problems with high dimensional feature space, common methods cannot easily yield satisfactory recognition results. Therefore, in these cases, we can employ machine learning techniques to identify the pattern of CSI changes and relate these to human behaviors. Machine learning includes numerous algorithms, such as DTW, HMM, Conditional restricted Boltzmann machines (CRBMs), SVM, KNN, Artificial Neural Network (ANN), BPNN, and SOM, etc. Although all these algorithms might generate similar classification results for the small data set or data with fewer features, they may produce a significantly different recognition result for large data set or data with a big number of features. Here we give the specific characteristics of machine learning algorithms utilized by the representative studies and applications.

a: DYNAMIC TIME WARPING

Dynamic Time Warping (DTW) is an important method that compares the similarity of two sequences, and it is also a template matching algorithm [191]. In order to judge similarity, DTW calculates the distance of the two sequences by utilizing a dynamic programming algorithm. There are some DTW classification applications of CSI, such as WiGeR [74] and Mudra [75]. WiGeR and Mudra utilize DTW to identify hand gestures. Even though DTW does not require training samples, it also has some side effects, such as a large amount of calculation and the recognition performance dependence on breakpoint detection and template, etc., [169].

b: SUPPORT VECTOR MACHINE

Support Vector Machine (SVM) is a common classifier in the pattern recognition field. It transforms the linearly indivisible samples of low-dimensional inputting space into high-dimensional feature space samples to divide the samples by using the nonlinear mapping algorithm, which realizes the sample separation in high-dimensional feature space. There are some SVM classification applications of CSI, such as DeNum [146] that achieves digit gesture recognition (e.g., 0 to 9) by using SVM. Although SVM has a better classification result, it is difficult to achieve multi-classification when the capacity sample is too large.

c: K-NEAREST NEIGHBOR

K-nearest neighbor (KNN) is a common classification method that classifies the data according to its K nearest neighbor. Specifically, KNN first finds the K nearest neighbor of the target and then calculates the node number of each type of K neighbors. We assign the target to the type based on its neighborhoods which hold the maximum nodes of that type. Currently, there have some human behavior recognition applications based on KNN. For example, FreeSense [115] applies KNN for user authentication and Wi-Wri [93] achieves the recognition of 26 English letters by using DTW and KNN. Despite the simplicity of theory and usability of deployment, KNN has some shortcomings as it just considers K nearest neighbors of one point. For data that have imbalance types or have large scale measurement values, the performance of KNN will severely decline.

d: HIDDEN MARKOV MODEL

Hidden Markov model (HMM) is a statistic model that usually is utilized to solve time sequence problems. Specifically, we can take the HMM to address the problems with the following two features. First, the problem describes sequence state changes, such as time sequences and state sequences. Second, the data of the problem consists of two types, observable data (observation sequences) and unobservable data (hidden sequences or state sequences). We infer the hidden sequence based on an observation sequence. Currently, several gesture recognition applications based on CSI have applied HMM to identify human movements. For example, Wang *et al.* [124] use HMMs to build a CSI behavior model for distinct motion states and realize the 8 kinds of behavior recognition. Despite the robustness for fine motion, HMM needs to tackle computation cost for the data with large hidden features.

D. MODEL-BASED HUMAN BEHAVIOR RECOGNITION

The purpose of model-based recognition is to leverage a mathematical or physical model to depict the signal variation caused by human behavior. Thereby, to identify human behavior, we have to build an appropriate model to interpret the CSI changes generated by user actions. However, many activities are conducted in indoor environments; therefore, the environment poses many effects on CSI propagation, which increases the difficulty of developing the physical model. Besides, CSI aims to evaluate wireless link state and improve communication qualification, which makes it difficult to correlate the specific signal change with the specific human behavior. In other words, the key of the model-based behavior recognition approach is to first establish the model that relates the signal space to the physical space and then utilize the model to determine the human behaviors based on the relationship between the received signal and the sensing target. Based on the physical model or mathematical model, human behavior can be accurately identified by exploring physical laws. In state-of-the-art CSI-based behavior recognition applications and studies, the typical models include Fresnel Zone Model, AoA, human respiration model, interacting model, CSI-speed model, and CSI-activity model.

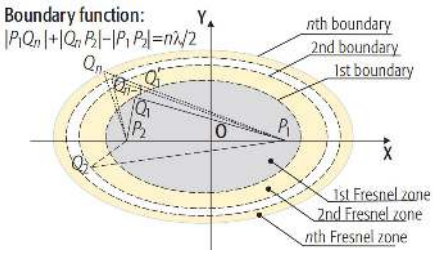

FIGURE 5. Fresnel zone model [186].

1) FRESNEL ZONE MODEL

The concept of Fresnel zone originated from the research on the interference and diffraction of light in the early 19th century. When applied in a radio propagation area, Fresnel zones refer to the series of concentric ellipsoids with two foci corresponding to the transmitter and receiver antennas [3]. As shown in Fig. 5, assuming P_1 and P_2 are two transceivers with a certain height, for a given radio signal with wavelength λ, the Fresnel zones containing n ellipses can be constructed by ensuring:

$$|P_1Q_n| + |P_2Q_n| - |P_1P_2| = n\frac{\lambda}{2} \tag{11}$$

where Q_n is a point on the n^{th} ellipse. The ellipse is the boundary of the Fresnel zone. Moreover, the innermost ellipse is defined as the 1st Fresnel zone, the elliptical annuli between the first and the second ellipse are defined as the 2^{nd} Fresnel zone, and the n^{th} Fresnel zone corresponds to the elliptical annuli between the $(n-1)^{th}$ and the n^{th} ellipses [192]. Apparently, the width of the Fresnel zone keeps decreasing as increasing of n and approaching $\lambda/2$.

From the mathematical view, Fresnel zone model describes the relationship between the geometric target position and the CSI power or amplitude caused by the target motion. In other words, when a target goes across different Fresnel areas, the signal path formed by the body reflection will vary with the different Fresnel area. Specifically, the different behaviors can be recognized by analyzing the signal path changes, such as walking direction estimation [129], respiratory detection [131]–[135], [193], human detection [126], and behavior recognition [125], [194].

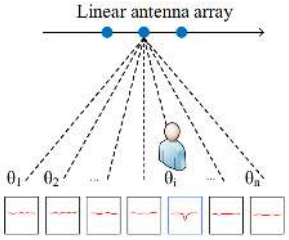

FIGURE 6. AoA estimation [176].

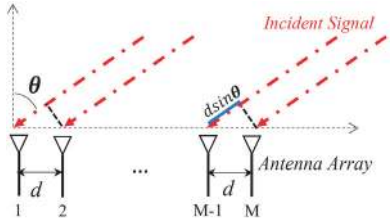

FIGURE 7. Antenna array with one incident signal.

2) ANGLE OF ARRIVAL

The angle of arrival (AoA) is a measurement that determines the propagation direction of RF waves on the antenna array. The basic idea is that whenever the user's body part occludes CSI measuring from a certain direction, the signal intensity of AoA corresponding to the same direction will descend distinctly [194]. Specifically, as shown in Fig. 6, the human body blocks the signal propagation along the AoA θ_i, but it does not affect signal transmission in other directions, i.e., θ_1, θ_2, θ_3, etc., which obviously leads to the weakness of the signal intensity of the AoA θ_i. In general, AoA can be evaluated by the phase difference between the antennas of the array [3]. As shown in Fig. 7, the phase difference between the adjacent antennas can be depicted as follows.

$$\varphi = e^{-j2\pi f\tau} = e^{-j2\pi fd\sin\theta/c} \tag{12}$$

where f is the frequency of the incident signal; c is the speed of the light; d is the interval between the adjacent antennas; $d\sin\theta$ refers to the path difference; $d\sin\theta/c$ refers to the time delay. Moreover, the more antennas, the better the AoA estimation performance. In order to acquire accurate angle estimation, the MUSIC (multiple signal classification) algorithm has been applied, which adopts two or more AoA

measurements from known points and utilizes triangulation to calculate the position of the signal source. All in all, the different behaviors can be tracked by analyzing the signal phase difference, such as handwriting recognition [130], human detection [126], [107], respiratory monitoring [136], and activity recognition [176].

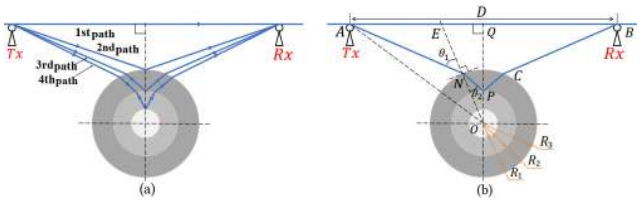

FIGURE 8. Interacting model [128]. (a) Signal propagation along multiple paths. (b) Detailed illustration of the 3rd path.

3) INTERACTING MODEL

Wang *et al.* [128] present an interacting model that realizes person identification. The model treats skin, fat, and bone as three concentric layers, and each layer's medium of signal propagation is uniform. Moreover, the user is asked to stand on the vertical line in the middle of the TX-RX antenna. As shown in Fig. 8 (a), the signal received S from the i^{th} path can be expressed as follows.

$$S = Ae^{-j2\pi f \sum_{i=1}^{n} (d_i \sqrt{\mu_i \varepsilon_i}) + \phi_0} \tag{13}$$

where A refers to the medium absorption of signal propagation process; ϕ_0 is the initial phase of the signal; d_i refers to the length of the i^{th} path; the permeability and permittivity are μ_i and ε_i, respectively. Wang *et al.* just consider the paths of three layers and signals reflected by the human body, as shown in Fig. 8 (b). The signal received S can be extended as:

$$S = Ae^{-j2\pi f \sum_{i=1}^{n} (h_i(\mu_0, \varepsilon_0, M, E, R, L) \sqrt{\mu_i \varepsilon_i}) + \phi_0} \tag{14}$$

where h_i refers to a function that calculates the propagation of the i^{th} path; the permeabilities, permittivity, and radiuses of the human body in each layer are M, E, and R, respectively; L means some location parameters.

4) CSI-SPEED MODEL AND CSI-ACTIVITY MODEL

Wang *et al.* [124] propose CARM in 2015, a human activity recognition system using the CSI-speed model and CSI-activity model [124], [186]. The principle of these two models can be demonstrated as follows. As shown in Fig. 9, when one person moves from P_1 to P_2, the WiFi signal is reflected by the wall and the human body, and propagation paths are changed, which causes multipath components. Specifically, the multipath components consist of LOS component and NLOS component caused by the wall, the human body, and other objects. The authors divide these multipath components into two types, such as static component (e.g., LOS component and the component caused by the wall and other objects) and dynamic component (e.g., the component caused by the human). Furthermore, the authors build

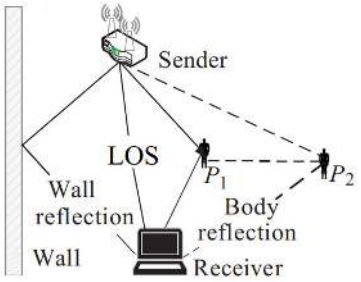

FIGURE 9. CSI-speed model [186].

the CSI-speed model and CSI-activity model to estimate the speed of human activities and multiple movement states by using the complex-valued channel frequency response (CFR) and analyze the amplitude and phase of CSI from static component and dynamic component. Due to some randomness of the CSI phase, it is difficult to develop a speed model by using phase information. Instead, the CSI-speed model builds a relationship between CSI amplitude and walking speed and evaluates the velocity of activities [186]. Furthermore, the author applies HMM to establish the CSI-activity model and identify human behaviors based on CSI energy changes on low-frequency components (e.g., caused by slow movement) and high-frequency components (e.g., caused by fast movement). CSI-activity model achieves great performance for quantifying the relationship between the velocity of human and activities.

E. DEEP LEARNING-BASED HUMAN BEHAVIOR RECOGNITION

Deep learning is one of the machine learning algorithms and can classify the data by exploiting the deep neural network (e.g., autoencoder, CNN, LSTM, and RBM, etc.). Usually, machine learning algorithms require accurate features as input because these features characterize the input data and determine output results. Therefore, well-designed features are the premise of correct behavior recognition and directly affect classification accuracy. However, some feature extraction may rely on empirical experience, decreasing classification accuracy. Different from machine learning, deep learning usually does not need feature extraction steps since it can automatically discovery and extract features from input data with a neural network model. Deep learning enables a new classification approach that can handle a large scale of data with complex features. In other words, with deep neural networks, there is no necessity to pre-process data for the acquirement of feature descriptions, which is the significant advantage of deep learning. Meanwhile, the large scale of unknown parameters in the neural network can automatically be calculated by the training process. Although the training process usually is extremely time-consuming, it can achieve satisfactory performance. Deep learning approaches have been widely adopted in many scenarios: from the classical image recognition applications to the challenging

natural language processing, up to visual art applications, and so on [195]. In this section, we analyze the specific neural network models applied in CSI-based behavior recognition applications, including autoencoder, CNN, LSTM, and RBM.

1) AUTOENCODER

Autoencoder is an unsupervised neural network model, which manages to rebuild the input data as the output by setting a narrow-hidden layer. Specifically, it can learn the implicit features of input data, which is called coding. Afterward, the original input data can be reconstructed with the new features learned, which is called decoding. Intuitively, it can be used for data dimensionality reduction because the neural network model can find hidden features that can reproduce the input data on the output by utilizing the internal representation layers. In addition, it can also be utilized as a feature extractor since the new features learned by the encoder can be fed into the supervised learning model. For example, Doong [158] use autoencoder to extract hierarchical features from the original input and achieve the estimation of people's numbers.

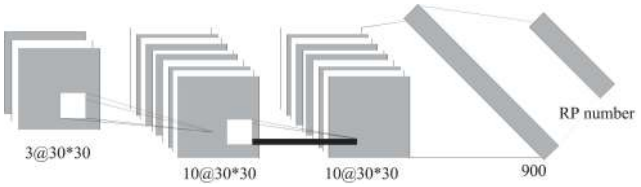

FIGURE 10. The structure of CNN.

2) CONVOLUTIONAL NEURAL NETWORK

Convolutional Neural Network (CNN) is a famous and representative deep learning model, which is used in image recognition and speech signal processing field extensively. As shown in Fig. 10, a typical convolution network consists of an input layer, some hidden layers, and an output layer with a full connection layer. Hidden layers comprise a sequence of convolution layers, rectified linear unit layers, and pooling layers. Specifically, the convolution layers are the crucial parts of hidden layers and implement convolution calculations. It can extract distinct features depicting the input data from a low level and concrete features to a high level and abstract features when hidden layers gradually increase. Relu layers finish nonlinear computation using activation function, which may build a complicated mapping relationship. The pooling layer usually includes maximization pooling or average pooling. Although pooling computation just chooses one from some results, it has become an indispensable part of deep learning as it can keep the translation invariance, rotation invariance, and scale invariance, decrease the number of parameters, reduce computation cost and alleviate overfitting. Jakkala *et al.*[152] use the deep convolutional network (DCNN) to identify the user's gait. DeepFalls [143] can automatically distinguish falls and fall-like activities by using DCNN.

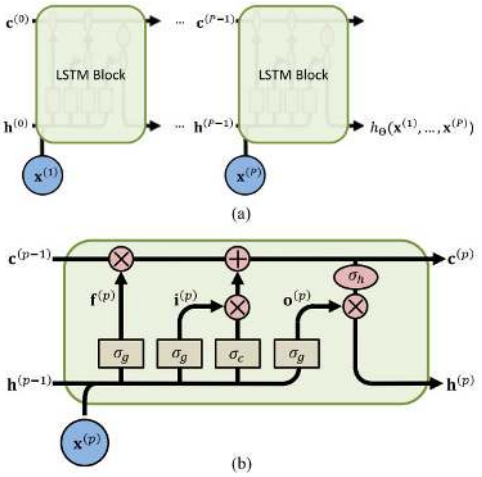

FIGURE 11. RNN structure with the LSTM block. (a) Overall structure. (b) The detailed structure of the LSTM block at the p^{th} time step.

3) LONG SHORT-TERM MEMORY NETWORK

The traditional neural network assumes that all inputs are independent of each other. This assumption may not be true for some applications such as natural language processing. Therefore, recurrent neural network (RNN) is proposed to address time sequences problems. Specifically, RNN is a simple cyclic neural network and can process sequential information. It performs the same task for each element of the sequence, whose outputs depend on the previous computation result. However, it cannot effectively tackle the long-term dependency problem because the output of the system depends on the previous information that occurred a long time ago. To solve this problem, some researchers propose the LSTM (see in Fig. 11). It is an improved RNN and can effectively process and predict important events with relatively long-term dependencies and intervals in time series. It has been used in various scenarios, such as image analysis, document summary, speech recognition, image recognition, handwriting recognition, music synthesis, and so on. Compared with RNN, LSTM adds a "processor" called cell to the algorithm to judge whether the information is useful or not. Three gates are placed in a cell, called the input gate, forget gate, and output gate. Information entering LSTM will be judged according to the specific rules. Only information certified by the algorithm will be retained, while other information will be forgotten through the forget gate. Based on this principle, LSTM is an effective technique for solving the problem of long-term dependency and achieves satisfactory performance. For example, Yousefi *et al.* [165] use LSTM to distinguish between lying down and falling. Feng *et al.* [178] recognize walking, running, and hand moving by applying LSTM, SVM, DTW with KNN.

4) RESTRICTED BOLTZMANN MACHINE

Restricted Boltzmann machine (RBM) is a type of improved Boltzmann machine (BM) to simplify network structure and

accelerate computation performance. BM, a random recurrent neural network based on the energy model, is proposed by G. E. Hinton *et al.* in 1984, to address complex learning problems by extracting the inherent representation of data. However, the learning processes of BM is complicated and time-consuming due to the connections among all nodes. To reduce computation complexity and enhance learning efficiency, RBM divides all nodes into two groups (e.g., input units and hidden units) and eliminates the link between units within the same group, which enables a more efficient training algorithm. RBM has been extensively applied in numerous scenarios, such as collaborative filtering, weight initialization, dimensionality reduction, classification, feature learning, topic model building, and deep belief network. In the case of human behavior recognition based on CSI, RBM can be used as feature extraction by using its powerful capability of representing complicated and hidden features. For example, Zhou *et al.* [146] built a stacked RBM for automatic feature extraction with higher accuracy and effectiveness to distinguish ten kinds of number gestures. Specifically, the authors first constructed the multi-layer RBM and then trained link weights with the probabilistic generative model layer by layer to obtain near-optimal initial values. Afterward, they unrolled the RBM and fine-tuned the weights with the back-propagation algorithm. Finally, the output of the last hidden units formed the effective features fed in the classification algorithm.

IV. CSI-BASED HUMAN BEHAVIOR RECOGNITION APPLICATIONS

Behavior recognition based on WiFi has received much research attention in recent years due to the popularity of WiFi devices indoors, the simplicity of approach, and the accuracy of the recognition results. In addition, since the CSI is sensitive to propagation path variation, it can be leveraged to depict these changes caused by human actions. As a result, CSI-based behavior identification techniques are widely used in the behavior recognition field. In general, human behavior recognition applications are categorized into two types: the pattern-based and the model-based applications based on whether the applications use the model to interpret human behavior. Although the recognition method based on deep learning can be regarded as a type of the pattern-based method, it is better to consider applications based on deep learning as a distinct recognition system due to its characteristics (e.g., without the feature extraction process and requiring a mass of data to train network parameters). Consequently, we categorize the state-of-the-art applications (as shown in Table 3) into three groups, including pattern-based, model-based, and deep learning-based. For better understanding and analysis of CSI-based human behavior recognition applications, we divide applications of each recognition technique into two types, such as specific behavior recognition and activity inference according to the purpose of behavior recognition.

Specific behavior recognition refers to some simple human behaviors which we just consider as distinct actions. Specifically, we only need to concentrate on the characterization of specific actions and can obtain their labels without further inferences. According to behavior granularities, we classify the specific behaviors into two types, such as coarse-grained specific behavior and fine-grained specific behavior. The coarse-grained specific behaviors refer to some activities which we conduct with a large range. Specifically, it consists of some simple daily activities and physical exercises, such as jogging, lying, sitting, cooking, washing, standing, walking, playing table tennis, bodyweight exercise, etc. These activities usually last some time with a certain periodicity and happen regularly. Besides, we regard some abnormal activities such as falling as coarse-grained behaviors, which usually occurs suddenly and experiences a very short time. The fine-grained behaviors refer to the specific human activities with a small motion distance. It consists of hand motion, lip movement, heart rate, and respiration, etc.

Differently, activity inference usually refers to determining the meaning of actions. Specifically, it aims to make the behavior inference to reveal the hidden purpose of the activities instead of specific behaviors. For example, when we want to count the number of persons in the room, we do not focus on someone's specific behaviors. We can count the number of crowds by identifying the differences between different individuals. The activity inference usually includes smoking detection, crowd counting, step counting, human detection, and user authentication, etc. Based on the analysis, we investigate the state-of-the-art applications and exhibit a comprehensive statistic based on the purpose of the applications, as shown in Table 3.

A. PATTERN-BASED BEHAVIOR RECOGNITION APPLICATIONS

Currently, there have plenty of pattern-based human behavior recognition applications using CSI because the pattern-based approach provides many advantages. According to identification purposes, the pattern-based applications are divided into three groups, such as coarse-grained specific behavior recognition applications, fine-grained specific behavior recognition applications, and activity inference applications. We make some tables to elaborate on the characteristics of typical applications from many aspects, including the device used, preprocessing methods, test scenarios, number of users, recognized behaviors, classifier, and recognition performance, as shown in Tables 4 to 6. We can obtain key features of each application from Tables 4 to 6 because these columns describe the characteristics of the applications. Besides, we categorize the applications based on the specific function, compare them from the above aspects, and obtain the key difference of these identification systems. In Table 4, there is one extra column compared with Table 5 and 6, titled TTW (Through the Wall), which is used to mark whether these applications can work in through-the-wall scenarios. The reason that the column solely exists in this table is that the

TABLE 3. CSI-based human behavior recognition applications.

Approach	Type	Human Behavior Recognition Applications
Pattern-based Behavior Recognition	Coarse-grained specific behavior recognition	Daily Behavior Recognition: E-Eyes [55], Wei *et al.* [56], WiHACS [57], MAIS [58], WiMU [59], DFS [60], TW-See [61], MultiTrack [62]. Table Tennis Action Recognition: WiTT [63]. Bodyweight Exercise Recognition: WiFit [64]. Danger-pose Detection: Zhang *et al.* [65]. Abnormal Activity Detection: NotiFi [66]. Falling Detection: WiFall [67], Anti-Fall [68], RT-Fall [69], FallDeFi [70], Dong *et al.* [71].
	Fine-grained specific behavior recognition	Hand Gesture Recognition: Nandakumar *et al.* [72], WiG [73], WIGeR [74], Mudra [75], WiFinger [76], WiFinger [77], PWiG [78], Chen *et al.* [79], WiCatch [80], iGest [81]. Sign Language Recognition: Shang *et al.* [82], WiSign [83]. Physiological Indicator Detection: Sleep Monitoring (Liu *et al.* [4], Wi-Sleep [84]); Respiration Detection (PhaseBeat [85], TensorBeat [86], TR-BREATH [87], Liu *et al.* [88]). Lip Reading and Speech Recognition: WiHear [89], WiTalk [90]. Keystroke Detection: WiKey [91], WindTalker [92]. Writing Recognition: Wi-Wri [93], LetFi [94], WriFi [95]. Sedentary Behavior Recognition: CareFi [96].
	Activity inference	Smoking Detection: Smokey [97]. Crowd Counting: FCC [98], FreeCount [99], Guo *et al.* [100], WiFree [101], Wi-Count [102], HFD [103]. Step Counting: Wi-Run [104], WiStep [105]. Human Presence Detection: DeMan [106], PeriFi [107], R-TTWD [108], SIED [109], PADS [110], Domenico *et al.* [111]. User Authentication: WiFi-ID [112], WiWho [113], WifiU [114], FreeSense [115], Wii [116], Chen *et al.* [117], AutoID [118], Wi-Sign [119], Liu *et al.* [120], BodyPIN [121], BioID [122], Liu *et al*. [123].
Model-based Behavior Recognition	Coarse-grained specific behavior recognition	Daily Activities Recognition: CARM [124], Zhang *et al.* [125].
	Fine-grained specific behavior recognition	Handwriting Recognition: WiDraw [130]. Respiration Detection: Wang *et al.* [131], TinySense [132], Demo [133], Yang *et al.* [134], Zhang *et al.* [135], FullBreathe [24], BreathTrack [136], FarSense [137].
	Activity inference	Human Presence Detection: FreeSense [126], RDFID [127]. User Authentication: WiPIN [128]. Walking Direction Estimation: WiDir [129].
Deep Learning-based Behavior Recognition	Coarse-grained specific behavior recognition	Daily Behavior Recognition: WiSDAR [138], DeepHare [139], DeepSense [140], Khan *et al.* [141], WiSPPN [142]. Falling Detection: DeepFalls [143], WmFall [144], Poster [145].
	Fine-grained specific behavior recognition	Hand Gesture Recognition: DeNum [146], FreeGesture [8], DeepNum [147], Widar3.0 [148], Temporal Unet [149]. Sign Language Recognition: SignFi [150]. Respiration Monitoring: Khan *et al.* [22].
	Activity inference	Gait Sensing and Walking Direction Estimation: Xu *et al.* [151], Jakkala *et al.* [152]. Crowd Counting: Chen *et al.* [156], WiCount [157], Doong *et al.* [158], DeepCount [159]. Human Presence Detection: CSI-Net [154], Wang *et al.* [153], Fang *et al*. [155]. User Authentication: Shi *et al.* [160], WiAU [161], NeuralWave [162], FingerPass [163].

coarse-grained behavior has a big activity range, which can be identified in through-the-wall scenarios. On the contrary, fine-grained behavior usually cannot be identified because the variation of CSI through the wall is too weak to be detected. Differently, behavior inference normally involves some continuous movements; therefore, we usually focus on the activities rather than the application scenarios. We analyze the key components of activity recognition applications as follows.

1) COARSE-GRAINED SPECIFIC BEHAVIOR RECOGNITION

a: DAILY BEHAVIOR RECOGNITION

Wu *et al.* propose a novel device-free through-wall human behavior identification system in 2018, called TW-See [61]. This system identifies seven kinds of behaviors (e.g., walking, hand swing, boxing, etc.) by applying the opposite robust PCA (Or-PCA), low-pass filter, and BP Neural Network. Specifically, the authors utilize low-pass filter to eliminate the noise of raw CSI data. Next, Or-PCA is applied to extract the correlation between human motion and changes of CSI values. Furthermore, the authors adopt a normalized variance sliding window algorithm to estimate the beginning and end of the behaviors to segment activity samples. Besides, eight kinds of features are fed into three-layer BP Neural Network, including STD, MAD, IR, DA, MEA, and so on. The authors collect samples in three different environments (e.g., environment 1 and 2: 12-inch concrete wall, environment 3: the top of the wall is glass, and the bottom is concrete.). The experiments demonstrate that TW-See can identify activities in a through-the-concrete-wall scenario with an average 94.46% accuracy.

TABLE 4. Pattern-based coarse-grained specific behavior recognition applications.

Type	Reference	Equipment	Sample	Preprocessing	Scenario	TTW	User	Behavior	Classifier	Performance
Daily behavior recognition	E-Eyes [55]	Intel 5300 NIC	400	Low-pass filtering, MCS index filter	two-bedroom 23ft×20ft, 24ft×36ft.	Yes	4	cooking, washing, etc., 9 behaviors.	DTW	TP: 96%, FP: 1%.
	Wei *et al.* [56]	WASP	4800	exponential smoother	living room, bedroom.	No	––	lying etc., 7 behaviors	SRC	Increasing 10% precision with RFI
	WiHACS [57]	Intel 5300 NIC	432	1-D linear interpolation, Hampel filter, DWT	3m × 8m, 6m × 8m, 9m × 8m.	Yes	––	Sitting, walking, falling, etc., 7 behaviors	SVM	LOS:97%, one-wall: 92%, and two-walls: 75%.
	MAIS [58]	Intel 5300 NIC	933	Low pass filter, Linear phase transform	Lab: 6m × 8m	Yes	––	sit, walk, fall, etc., 7 behaviors.	KNN	LOS:97%, one-wall: 92%, two-wall: 75%.
	WiMU [59]	Intel 5300 NIC	1050	PCA	Lab: 25ft × 16ft, 6ft × 3ft, 6ft × 6ft.	No	10	Open/close door, etc., 6 actions.	CFR	Na:95%, 94.6%, 93.6%, 92.6%, 90.9%.
	DFS [60]	Intel 5300 NIC	–––	median filter, FFT	11.7×8.4m^2, 0.8× 7.8m^2, corridor	No	2	8 actions: bow, swing arms, etc.	SOM neural network	Best: DFS-CHB: 0.879.
	TW-See [61]	Intel 5300 NIC	2240	low-pass filter, PCA	office: 5m × 8.7m, meeting: 12m × 8.7m.	Yes	6	walk, fall, sit, etc., 6 behaviors.	BP-network	Average: 94.46%.
	MultiTrack [62]	Intel 5300 NIC	9000	NDFT	home:15ft × 20ft, corridor, classroom: 25ft×30ft.	No	5	6 actions: squat, sit up, lunge, etc.	MD-DTW	over 92%
Table tennis action recognition	WiTT [63]	Intel 5300 NIC	1350	DWT, PCA, Butterworth low-pass filter	living room, apartment.	No	––	forehand attack, etc., 9 activities.	SVM, DTW+KNN	Detecting: 96.34%; 6 actions: 90.33%.
Bodyweight exercise recognition	WiFit [64]	Intel 5300 NIC	720	none	office: 4m × 4m, meeting room: 6m × 5m.	No	20	bodyweight exercises: sit-up, etc	Doppler-MUSIC, SVM	Repetition counting: 99%, Exercise type: 95.8%.
Danger-pose detection	Zhang *et al.* [65]	Intel 5300 NIC	1615	Phase Calibration	bathroom	No	––	Danger activities when takes shower	SVM	Precision: 83.61%. Recall: 96.23% F1: 89.47%.
Abnormal Activity Detection	NotiFi [66]	Intel 5300 NIC	180	PCA	office: 9m × 9m, Lab: 9m ×10m, Apart: 8×9m^2.	Yes	7	slip, fall, run, breath pausing.	Bayesian model	LOS :89.2%, NLOS: 85.6%, one-wall: 75.3%.
Falling Behavior Detection	WiFall [67]	Intel 5300 NIC	150	anomaly detection algorithm LOF	chamber: 3m × 4m, Lab: 8m × 9m, dorm room: 4m × 5m.	No	10	walk, fall, etc., 5 behaviors	SVM Random Forest algorithm	87% detection precision with a false alarm rate of 18%.
	Anti-Fall [68]	Intel 5300 NIC	740	Interpolation, low-pass filtering	office: 3m × 4m, meeting: 6m × 6m.	No	6	sitting, fall, etc., 9 actions	SVM	89% detection with 13% of false alarms.
	RT-Fall [69]	Intel 5300 NIC	Each 200	low-pass filtering	office: 3m × 4m, Apart, meeting: 6m × 6m, hall: 12m × 6m.	No	6	walk, sit, lie, stand, squatting	SVM	RT-Fall is 14% and 10% higher than WiFall.
	FallDeFi [70]	Intel 5300 NIC	A:571; B:184	interpolation, DWT, PCA	Lab: 7.3m × 4.6m, Apart : 4.2m × 15m, bathroom: 2.9 m × 4m.	No	3	5 different types of falls	SVM	Average:93 %, Improve: RT-Fall-12 %, CARM (e.g., falling)-15 %.
	Dong *et al.* [71]	Intel 5300 NIC	4800	Hampel and low-pass filter	corridor, room, staircase.	No	8	fall, run, walk, and sitting.	SVM, KNN	falling: flat floor-92%, staircase-94%.

TABLE 5. Pattern-based fine-grained specific behavior recognition applications.

Type	Reference	Equipment	Sample	Preprocessing	Scenario	User	Behavior	Classifier	Performance
Hand Gesture Recognition	Nandakumar *et al.* [72]	Intel 5300 NIC	——	low-pass filter	office	6	push, pull, punch, etc.	Peak detection	91%
	WiG [73]	Intel 5300 NIC	Each 180	Birge-Massart filter	single room: 5 m ×6 m	—	4: waving, etc.	SVM	LOS: 92%, NLOS 88%.
	WIGeR [74]	Intel 5300 NIC	300	Butterworth low-pass filter	laboratory: 8m × 8m	6	swiping etc, 7 hand gestures	DTW	Env 1,2,3,4: 97.28%, 91.8%, 95.5%, 83.85%.
	Mudra [75]	NI-based SDR	——	neighboring range averaging	office: 5m × 7m	5	9: shoot, pick, come, etc.	linear regression, DTW	96%
	WiFinger [76]	Intel 5300 NIC	315	Butterworth low-pass filter, DWT.	office	10	9 ASL gestures.	DTW, KNN	Average: 90.4%.
	WiFinger [77]	Intel 5300 NIC	——	DWT, Bandpass filter	home: 16ft × 13ft, office: 9ft × 9ft.	>2	zoom out and in, etc. 8 gestures	MD-DTW	93%
	PWiG [78]	Intel 5300 NIC	——	Heursure threshold wavelet	6m × 5m	—	5: swipe left, etc.	KNN, SVM	LOS: 94%, NLOS: 89%.
	Chen *et al.* [79]	Intel 5300 NIC	115	PCA, Low pass filter, DWT.	indoor, experiment	5	5 gestures.	DTW, KNN.	95%
	WiCatch [80]	Intel 5300 NIC	——	Phase Sanitization	33 m^2	—	9: boxing, etc.	SVM, GDC	over 0.96, two-hands: 0.95
	iGest [81]	Intel 5300 NIC	480	Savitzky-Golay filter	Lab, class, meeting	4	4: push, etc.	HMM	Best: 91%
Sign Language Recognition	Shang *et al.* [82]	Intel 5300 NIC	1600	median filter, low-pass filter.	indoor	3	8: hello, no, etc.	SVM	accuracy/FP: 93.8%,1.55%.
	WiSign [83]	Intel 5300 NIC	Train: 6 × 15	spikes removal /low-pass filter	indoor	6	8: Hello, no, etc.	SVM.	87.01%, 87.38%.
Physiological Indicator Detection	Liu *et al.* [4]	Intel 5300 NIC	——	Hampel filter, average filter.	Lab: 12ft×9ft, 15ft×12ft.	2	sleeping posture: curl up, etc.	PSD, K-means clustering	postures: 80%, Prone: 93%.
	Wi-Sleep [84]	Intel 5300 NIC	——	Hampel filter, wavelet filter.	bedroom, meeting	—	6 sleeping positions.	SVD	detection error: less than 20%.
	PhaseBeat [85]	Intel 5300 NIC	——	Hampel filter, DWT	Lab: 4.5m × 8.8m, 20m-corridor	—	breathing, heartbeat	root-MUSIC, FFT-based	Heart: error medium-1bpm, breath: 98%, 2 users-90%
	TensorBeat [86]	Intel 5300 NIC	——	autocorrelation	Lab: 4.5m × 8.8m, corridor:20m	—	breathing,	DTW, signal matching	1 user:96%; 2 or 3 users:93%; 5 users:62%.
	TR-BREATH [87]	Intel 5300 NIC	84	Phase calibration	5.5m × 5m, 8m × 7m, 8m × 5m	9	multi-person breathing rate:5~11m.	MUSIC, SVM.	1-user-NLOS: 99%; dozen-LOS/NLOS: 0.9865, 8.07%
	Liu *et al.* [88]	Intel 5300 NIC	——	Data calibration, moving average filter, PCA.	two apartments: 12ft × 9ft, 15ft × 12ft.	2	sleep postures, heart rate	PSD, SVM, KNN, RF	postures: over 90%; heart rate errors: less 2b/min-0.57, 4b/min-0.9
Lip Reading and Speech Recognition	WiHear [89]	USRP N210	50~100	Butterworth IIR band-pass filter.	Hall: 9m × 16m	4	9 types of vowel sounds.	DTW, MCFS, SRDA.	3-users: 74%, 1-user: 91%.
	WiTalk [90]	Intel 5300 NIC	——	DWT, PCA, Band pass filter	5.7m × 6m	3	12 elements	CFR, DTW	92.3%, 6 words: 74.3%
Keystroke Detection	WiKey [91]	Intel 5300 NIC	train:30 ~80 × 37	Butterworth low-pass filter, PCA	library	10	fingers	KNN	1-key: 96.4%, sentence: 93.5%.
	WindTalker [92]	Intel 5300 NIC	1200 keys	Low pass filter, PCA	AP-PC: 75 cm	10	fingers and direction	DTW	79.0%, 63.0%
Writing Recognition	Wi-Wri [93]	Intel 5300 NIC	(24~48) × 26	Butterworth filter, PCA	laboratory	8	written letters	DTW, KNN.	82.7%
	LetFi [94]	Intel 5300 NIC	32 × 26	PCA	experiment	—	26 capital letters.	SOM	95%
	WriFi [95]	Intel 5300 NIC	——	PCA, FFT	Lab: AP-PC:1.6m	5	26 letters.	HMM	88.74%
Sedentary Behavior Monitoring	CareFi [96]	Intel 5300 NIC	——	PCA, FFT	dorm room office	5	sitting still, drink, etc.	PPMCC	Average: 94.9%

TABLE 6. Pattern-based activity inference applications.

Types	Reference	Equipment	Sample	Preprocessing	Scenario	User	Behavior	Classifier	Performance
Smoking Detection	Smokey [97]	Intel 5300 NIC	1918	Data Interpolation	7.5m × 4.5m 4.8m × 3.6m	2	6 smoke actions	autocorrelation function	92.8%
Crowd Counting	FCC [98]	Intel 5300 NIC	———	PEM	Lab, meeting room, office, outdoor.	15	moving	Gray Verhulst model	1-user (indoor-98%), 2-users (outdoor-70%)
	FreeCount [99]	Atheros	7000	Wavelet denoising	conference: 4.5m × 6.5m	7	walking casually	TKL, SVM	96%.
	Guo *et al.* [100]	Intel 5300 NIC	286	THD, FFT.	7 × 6m^2, 12 × 5m^2, 16 × 10m^2, 13 × 7m^2, 12 × 5m^2.	10	walking speed, direction	semi-supervised learning	More than 90%
	WiFree [101]	Atheros	———	DWT	2.6m × 5.2m, conference: 6.8m × 5.1m, 7.6m × 7.4m.	7	walking	SVM, TKL	crowd counting: 92.8%.
	Wi-Count [102]	Intel 5300 NIC	———	PCA	office, doorway	8	walk	ICA, source separation	direction: 95%, counting: 92%.
	HFD [103]	Intel 5300 NIC		window-based filter	conference: 8.7m × 5.8m, corridor.	——	walk	SVM, RBF	90%
Step Counting	Wi-Run [104]	Intel 5300 NIC	———	Data calibration Hampel filter.	15.4m × 8.7m, 11 treadmills.	5	run	CP decomposition	88.25%
	WiStep [105]	Intel 5300 NIC	8400 + 5600	PCA, DWT, bandpass filter,	7m × 8.5m, 8m × 6m	15	walking	Short-time energy	Lab: 90.2%, class room: 87.59%.
Human Presence Detection	DeMan [106]	Intel 5300 NIC	———	bandpass filter	class room, Lab	——	Moving, static	PSD, SVM	Moving:90.93% Breath:90.44%
	PeriFi [107]	Intel 5300 NIC	———	sanitize phase	9m × 6m.	——	walk, stationary	SVM, LOSO, AoA	96.7%
	R-TTWD [108]	Intel 5300 NIC	4320	low-pass filter	meeting room: 9m × 6m, office: 5m × 9m.	4	walking speed	SVM	99%
	SIED [109]	Intel 5300 NIC	———	Window function	meeting room, Lab: 10m × 6m	——	walking	HMM	9m: 94%
	PADS [110]	Intel 5300 NIC	———	Phase sanitization	corridor, Lab	——	walking	SVM	1.2m to 2m: 98%
	Domenico *et al.* [111]	Intel 5300 NIC	———	FFT	5m × 3.5m, 5m × 9m, 5m × 6m	——	walking	Naïve Bayes	Through-one-Wall: 99%, 91%
User Authentication	WiFi-ID [112]	Intel 5300 NIC	———	Butterworth filter	corridor	6	walking, body	SAC, CWT	2 users: 93%, 6 users: 77%.
	WiWho [113]	Intel 5300 NIC	540	Butterworth band-pass filter	office, residential.	20	gait	DTW	2~6 persons: 92%~80%.
	WifiU [114]	Intel 5300 NIC	2800	PCA, STFT.	7.7m × 6.5m	50	gait	SVM	Top 1,2,3: 0.79, 0.89, 0.93.
	FreeSense [115]	Intel 5300 NIC	360	low-pass filter, PCA, DWT.	home: 6m × 5m	9	walking	DTW, KNN	2~6 people: 94.5%~88.9%.
	Wii [116]	Intel 5300 NIC	≥1500	PCA, CWT, Low-pass filter.	meeting room:5m × 4m	8	gait	SVM, GMM	2 users: 98.7%, 8 users: 90.9%.
	Chen *et al.* [117]	Intel 5300 NIC	1800	Low-pass filter	Lab: 9m × 6m; corridor: 2m × 7.5m; meeting: 8m ×5m	20	gait	LibSVM with RBF	2~6 users：92% ~ 82%
	AutoID [118]	Atheros	400	———	meeting: 5m × 7m, office: 5.6m × 9m, apartment: 7.5m × 8m.	20	gait	C^{3}SL, AGGM	91%

TABLE 6. ***(Continued.)*** **Pattern-based activity inference applications.**

	Wi-Sign [119]	Intel 5300 NIC	112	Butterworth low pass filter, PCA, DWT.	office: 5 × 3.5m	14	hand motion	LIBSVM	14 users-TPR: 79%, attacks-TNR: 86%.
	Liu *et al.* [120]	Intel 5300 NIC	99	K-means algorithm	Lab: 11m × 12m, Apart: 11m × 6m	2	moves	SVM	Best: Sce-1 DR: 100%, FPR: 0.1%
	BodyPIN [121]	Intel 5300 NIC	Test : 600	Butterworth low-pass filter.	5m × 6m	30	sitting	LIBSVM	30 users: 92%.
	BioID [122]	Intel 5300 NIC	300	Butterworth low-pass filter, PCA, DWT.	office	5	lip-motion	DTW, KNN, Decision Tree.	90%
	Liu *et al.* [123]	Intel 5300 NIC	–––	EMD-based filter	office:17ft× 9ft	20	keystrokes, gaits	KNN	over 93%

b: TABLE TENNIS ACTION RECOGNITION

Chen *et al.* propose WiTT [63] in 2018, a unique system that can detect and classify table tennis activities by analyzing the changes of CSI values. Firstly, the authors use DWT, PCA, and Butterworth low-pass filter to wipe out the noise and extract useful features. Next, the authors utilize SVM to train samples and classify 9 activities of playing table tennis, such as forehand attack, backhand stroke, forehand pull, forehand pick, backhand pull, backhand close shot, step, step by step, and squat. The extensive experimental results prove that WiTT only has an average 79.78% accuracy for 9 kinds of behavior identification due to some similar actions, 90.33% accuracy for six types of activity recognition, and more than 96.34% detection rate for presence detection of table tennis actions.

c: BODYWEIGHT EXERCISE RECOGNITION

Li *et al.* propose a device-free bodyweight exercise monitoring system in 2018, called WiFit [64]. This system can accurately count the repetition of bodyweight exercise by analyzing the Doppler effect caused by human activity. This system first applies the impulse-based method to automatically segment each exercise, and then SVM is utilized to count the number of activities according to doppler features of three exercises (e.g., sit-up, push-up, squat). The authors invite 20 persons to participate in experiments and collect a total of 4350 samples in the office and meeting room. The results show that this system can count the number of repetitions of bodyweight exercise with an accuracy of 99% and accurately classify exercise type with an accuracy of 95.8%.

d: DANGER-POSE DETECTION

Zhang *et al.*[65] propose a device-free danger-pose detection system in 2018. This system can automatically identify three kinds of dangerous behaviors when taking a shower in the bathroom, including steady lying position, the whole-body sinks below the water surface, and the face sinks below the water surface. The central components of this system consist of preprocessing, activity classification, and danger detection. Specifically, the authors select amplitude and phase as base signals and then calibrate phase. Furthermore, the authors extract static and dynamic features from CSI low-frequency signals and feed these features in one-class SVM. The authors evaluate the performance of this system using three types of criteria, such as precision, recall, and F1-score. According to experimental results, it demonstrates that this system accurately recognizes danger-pose with 83.61% precision, 96.23% recall, and 89.47% F1-score.

e: ABNORMAL ACTIVITY DETECTION

Zhu *et al.* propose a device-free non-invasive abnormal activity detection system in 2017, called NotiFi [66]. This system leverages some steps including PCA, non-parametric Bayesian model, and Hierarchical Dirichlet Process (HDP) to detect five kinds of abnormal actions (e.g., slipping on the ground, falling, running, and breath pausing, etc.). Specifically, the authors first exploit PCA to remove noise and then design a model to describe the activity as the CSI state trajectory, including amplitude and phase information. The abnormal activities can be identified by exploiting the non-parametric Bayesian model and Dynamic Hierarchical Dirichlet Process. NotiFi is evaluated in three scenarios (e.g., LOS, NLOS, through-one-wall scenario) from three rooms, including office, laboratory, and apartment. The experimental results show that NotiFi has a fine performance with an accuracy of 89.2% in LOS, 85.6% in NLOS, and 75.3% in a through-one-wall scenario.

f: FALLING DETECTION

Wang *et al.* propose a real-time, contactless, and cost-saving indoor fall identification system, called RT-Fall [69] in 2017. The authors utilize the phase, amplitude, and phase difference of CSI to detect falling motion. The authors discover that the phase difference of CSI is sensitive to fall and fall-like activity. Therefore, the authors exploit the phase difference for activity segmentation based on the correlation between human activities and phase change variance. In addition, the end of fall and fall-like activity can lead to a sharp power profile decline pattern in the time-frequency domain, which can be utilized to validate the activity segmentation. After signal preprocessing (e.g., interpolation and band-pass filter), the fall-like activity segmentation is performed to separate

fall-related activities. Next, eight features are obtained in real-time and fed into v-SVM to identify fall activity. Experiments are conducted in four indoor scenarios, and results are compared with the WiFall to validate the algorithm performance. The results demonstrate that RT-Fall can separate the falls with the precision 100% and achieves 91% of sensitivity and 92% of specificity. Moreover, the sensitivity and specificity on average of RT-Fall are 14% and 10% higher than that of WiFall.

2) FINE-GRAINED SPECIFIC BEHAVIOR RECOGNITION

a: HAND GESTURE RECOGNITION

Tian *et al.* propose a specific device-free hand gesture identification system, called WiCatch [80] in 2018. This system identifies 9 kinds of hand gestures by applying phase information of CSI, including one-hand gestures (e.g., pull, push, slide, leftward, rightward, and wave hand) and two-hand gestures (e.g., boxing, open the fridge, and open the window). This system consists of four main processes, such as interference elimination, virtual array construction, motion trajectory reconstruction, and gesture classification. Specifically, the authors eliminate phase errors caused by SFO, CFO and STO (a random time shift that can cause phase offset of CSI) and then utilize the Gerschgorin Disk Criterion (GDC) algorithm to estimate the number of angles of hand gestures. Moreover, the authors apply SVM and radial basis function (RBF) kernel to classify the spectrums of hand gesture and achieve single-hand and two-hand gestures identification with 0.97 and 0.96 average accuracies, respectively.

b: PHYSIOLOGICAL INDICATOR DETECTION

Liu *et al*. [88] propose a breathing and heart rate detection system during human sleeping in 2018, which can identify sleep postures (e.g., curl up, supine, prone, and recumbent) and realize breathing rate tracking. The main components of the system consist of CSI collection, coarse sleep event detection, heart rate estimation, breath rate estimation, and sleep event and posture identification. Specifically, the authors apply Hampel filter to remove outliers of each CSI subcarriers, and then a moving average filter is adopted to remove high-frequency noise that does not come from breath breathing or heartbeats. After that, the authors apply the PSD-based approach to estimate the human breathing rate. To evaluate the system performance, the authors utilize another four different techniques to identify respiration rates, such as discriminant analysis (DA), KNN, SVM, and RF. This system can infer heart rate with errors of 57% are less than 2 b/min and over errors of 90% are less than 4 b/min. Moreover, this system achieves over 90% accuracy for four sleep postures recognition using KNN ($K = 5$), SVM, and RF.

c: SIGN LANGUAGE RECOGNITION

Shang *et al.* propose WiSign [83] and improve it [82] in 2017, two sign language recognition systems that recognize eight sign languages such as Hello, Thanks, Yes, No, etc. The main characteristics of WiSign are that this system leverages training data that solely have sparse labels to recognize gestures by using transfer learning and semi-supervised learning. Specifically, the authors choose two features as input vectors from the calculated eight waveform statistic features based on the received raw signal. Next, the authors classify unlabeled data with small labeled data based on SVM and KNN to use the semi-supervised learning-based solution. In addition, the authors discrete the feature values of any two samples and judge whether they are similar according to the absolute value of discretization feature calculation formulation. Based on this calculation and SVM, the authors implement the transfer learning-based solution. The experiments show that WiSign can achieve the mean prediction accuracies of 87.01% with the transfer learning-based approach and 87.38% with the semi-supervised learning-based approach for all participants, which are better than traditional SVM. In order to improve recognition accuracy, another laptop is introduced and different kernel functions of SVM is applied [82]. Meanwhile, for some complicated gestures whose signal often overlaps, a weighted voting system is employed. The experiments show that the system obtains a better mean false positive of 1.55%, and enhances the recognition accuracy to 93.8% compared with having only one laptop in the same environment.

d: LIP READING AND SPEECH RECOGNITION

Wang *et al.* propose WiHear [89] in 2016, a novel system to hear talks with WiFi signals. Two core components include the mouth motion profile for feature extraction and the learning-based signal processing approach for lip reading. Mouth Motion Profile construction consists of the following steps: localizing the mouth, filtering out interference and reflection using 3-order Butterworth IIR band-pass filter, partial multipath removal using delay over 500 ns in the time domain and FFT, and building mouth motion profile using the calculation of a coefficient C. The coefficient C depicts the peak to peak value on every subcarrier in a sliding window and is representative value of every time slot. Afterward, authors apply discrete wavelet packet decomposition of 4-order Symlet wavelet filter on the filtered signal. The lip-reading includes the following steps: segmentation with detection of the silent interval, feature extraction with Multi-Cluster/Class Feature Selection (MCFS) scheme that extracts main features from wavelet profiles, classification with dynamic time warping (DTW). The authors conduct experiments for recognition of 14 different syllables under six scenarios, including LOS, NLOS, and through-the-wall environments. Experiment results show that WiHear can achieve recognition accuracy of 91% for a single person speaking less than seven words and up to 74% for hearing less than four users simultaneously.

e: KEYSTROKE DETECTION

Ali *et al.* propose a keystroke recognition system in 2015, called WiKey [91]. This system can detect keystroke accord-

ing to the movement of hands and fingers. Firstly, the authors apply Butterworth filter and PCA to eliminate some high-frequency noise and extract the signals which only contain variations caused by movements of hands. The authors determine the start and end of activities by detecting an increase and decrease in rates of change in the CSI time series. Based on the experiment observation, means and variances or frequency components calculated from waveform cannot be used as feasible features due to the similarity between different keystrokes. Therefore, the authors utilize the extracted keystroke waveforms as feature representation since the shapes containing both time and frequency domain information of the waveforms are suitable for classification. The authors choose Daubechies D4 (four coefficients per filter) wavelet and scaling filters because they preserve enough information and finish the maximum computation. WiKey utilizes an ensemble of KNN using DTW distance as the comparison metric between keystroke shape features. Extensive experiments demonstrate that WiKey achieves more than 97.5% keystroke detection rate and 96.4% single keys recognition accuracy. In real-world experiments, WiKey achieves a recognition accuracy of 93.5% for keystrokes in a continuously typed sentence.

f: WRITING RECOGNITION

Zhang *et al.* propose a novel device-free letter recognition system, called LetFi [94] in 2018. This system can successfully detect the hand motion of writing 26 capital letters in the air and estimate which letter has been written. The main processes consist of robust CSI measurement extraction, the detection of the start and end point of the letter-writing activities, multi-domain feature extraction based on coherence histogram, and action classification using SOM network. Specifically, the authors apply PCA to extract the 1st principal component and utilizes it as the metric to detect the start or end of the letter-writing action. Furthermore, the authors utilize Fast Fourier Transformation (FFT) to acquire multi-domain features that contain amplitude and phase information of CSI matrixes. Next, the authors use SOM neural networks to train and test 832 ($16 \times 26 \times 2$) CSI samples and realize accurate classification of multi-domain features. The results show that LetFi has 95% identification precision for the 26 capital letter recognition.

g: SEDENTARY BEHAVIOR MONITORING

Yang *et al.* propose CareFi [96] in 2018, a device-free Sedentary Behavior (SB) monitoring system. In CareFi, the authors categorize sedentary behaviors into dynamic and static activities which have different properties. The authors find that different subcarriers are sensitive to different parts of the human body. They employ PCA to obtain the trends of CSI changes and distinguish dynamic activities based on the main features such as variance or kurtosis. Different from other applications, the authors utilize the frequency information to capture the critical features representing the static postures and occupancy position. In addition, they propose a foreground detection method based on kernel density estimation (KDE) to extract coarse motion. On the preprocessing phases, linear interpolation, subcarrier refinement with IFFT and FFT, and DWT coefficients for temporal de-noising are utilized to acquire the sanitized CSI data. For dynamic activities, the authors choose the first four PCA components and eight features to describe activities, including entropy, mean, variance, standard deviation, kurtosis, skewness, interquartile range, and mean crossing rate. The authors apply Pearson product-moment correlation coefficient (PPMCC) to measure the similarity between two signal vectors and distinguish static postures. According to the experiment results, CareFi achieves 94.9% average accuracy for six common activity recognition, including reading, write, type, drink, sit up or down, and phone. The classification accuracy of human activities under environmental interference achieves 97.5%, and accuracy for other conditions vary from 90.5% to 100.0%.

3) ACTIVITY INFERENCE

a: SMOKING DETECTION

Zheng *et al.* propose Smokey [97] in 2016, a ubiquitous passive smoking detection system. The smoking procedures are decomposed into six steps in a certain order. The authors find that different smoking phases affect different subcarriers because the subcarriers have distinct sensitivity to the motions of different parts of the human body. Based on this principle, the authors utilize the rhythm/order information to detect smoking activities from time-varying and subcarrier-dependent CSI information. The system consists of data processing, motion acquisition, and activity analysis. Specifically, due to the inherent noise of the CSI stream, linear interpolation is adopted on the irregular data to obtain the CSI sequence with samples evenly spaced in time. Afterward, the variations of CSI caused by motions are detected from dynamic noises utilizing the image processing algorithm, and the composite motions are acquired based on the temporal correlation and the frequency correlation. Then, the authors utilize autocorrelation to analyze the periodicity of the composite motions in each detection window and identify smoking motions according to the range of smoking period and the threshold of the standard deviation of periods. This system achieves 92.8% detection accuracy for the smoking activities and misjudges normal activities as smoking action with 2.3% error. In a static environment, the true positive rate (TPR) of Smokey can be as high as 0.976, and the average false positive rate (FPR) is low to 0.008, while in a dynamic environment, the TPR reduces to 0.919 and the FPR increases to 0.097.

b: CROWD COUNTING

Zou *et al.* propose FreeCount [99] in 2017, a device-free crowd counting system based on CSI data. Firstly, the authors apply the wavelet-based de-noising scheme to remove the ambient noise of the raw CSI data. Then, some features

(e.g., statistics features, transformation-based features, and shape-based features) are calculated. Afterward, the authors choose the most representative features which are sensitive to human motion while resilient to environmental variation. Finally, rather than using the raw CSI feature space, the authors develop a robust crowd counting classifier based on transfer kernel learning (TKL) in the reproducing kernel Hilbert space to construct a domain-invariant kernel for SVM training. This system with TKL achieves 91.97% estimation accuracy, outperforming SVM with RBF kernel by 38.78%. In all, FreeCount estimates the number of people with 96% accuracy up to 7 users, consistent with that under temporal and environmental variation.

Zou *et al.* propose WiFree [101] in 2018, a novel application for occupancy detection and crowd counting. This system applies OpenWrt and the Atheros CSI tool, which collects more CSI data related to human behavior. The authors apply the low-pass filter and wavelet-based de-noising scheme to filter the noise of CSI data and then sanitize the raw CSI stream. In addition, the authors apply DTW to extract features and count the crowd by using transfer kernel learning (TKL) and SVM. This system detects seven volunteers in three rooms (e.g., discussion room, conference room, and seminar room) of different sizes. The results demonstrate that WiFree achieves 99.1% accuracy for occupancy detection and 92.8% accuracy for crowd counting.

c: STEP COUNTING

Zhang *et al.* propose a device-free multi-runner step counting system in 2018, called Wi-Run [104]. This system can automatically estimate the number of steps when users run on the treadmill by three novel approaches, including Canonical Polyadic (CP) decomposition, stable signal matching, and the peak detection method. Specifically, the authors apply Hampel filter and Savitzky-Golay filter to eliminate noises that contain both low and high-frequency components. Next, Canonical Polyadic is applied to decompose signals related to running. Finally, the authors apply a peak detection method to estimate steps for each runner. The experiment results show that Wi-Run has 88.25% average accuracy for overall step estimation. Moreover, this system has an average of 91.30%, 90.21%, 86.97%, 84.53% accuracy when runners increase from two to five, respectively.

Xu *et al.* propose a novel device-free step counting system based on the torso frequency analysis, called WiStep [105] in 2018. This system can count in-place walking steps even when the user's torso speed is almost zero. Firstly, the authors apply Butterworth bandpass filter, PCA, and DWT to filter high-frequency noise caused by external interference, and then the time-frequency analysis method is utilized to segment and identify the walking by analyzing amplitude variances of CSI streams. Furthermore, wavelet decomposition is applied to extract useful coeffcients which are related to these frequencies induced by feet or legs. Finally, the authors apply the Short-time energy of the coeffcients to count steps. According to extensive experiments, WiStep achieves 90.2% and 87.59% accuracy for overall step counting in laboratory and classroom, respectively.

d: HUMAN DETECTION

Zhu *et al.* propose R-TTWD [108] in 2017, a novel device-free moving human detection system in a through-the-wall environment. This system employs Hampel identifier, PCA, low-pass filter, and SVM to detect human behavior. The authors adopt the Hampel filter to remove outliers that are some points falling out of the closed interval, and then they apply the 1-D linear interpolation algorithm to obtain continuous samples located in consecutive and even time slots. Afterward, they utilize low-pass filter and PCA to remove noise at the top position of the spectrum. After human detection and feature extraction, the authors apply SVM to classify the samples. The system achieves over 99% detection rates in many general experimental scenes which have different wall materials, dynamic moving speeds, and so on.

e: USER AUTHENTICATION

Shah *et al.* propose Wi-Sign [119] in 2018, a unique two-factor authentication (2FA) system that does not depend on an auxiliary device to determine the second factor. This system recognizes user identity by exploiting the signing motion. In other words, the authors identify hand geometry and the way of hand movement to estimate the people's identity. Firstly, the authors utilize the Inverse Fast Fourier Transform (IFFT) to transform the frequency domain of CSI to time-domain Channel Impulse Response (CIR), which deletes more than 500ns multipath delay. Next, the authors apply low pass Butterworth filter, DWT, and PCA to filter high-frequency noise and acquire detail-coefficients. A threshold method based on the first-order variance of the selected principal component is applied to determine the start and end of hand motion. Besides, the authors use one-class SVM (LIBSVM) with RBF kernel to classify eight features, including mean, skewness, kurtosis, standard deviation, etc. The experiment results show that Wi-Sign has an average 79% TPR for user's identity recognition of 14 users and has an average 86% TNR for attack detection (The common attack is that the intruder mimics the signer).

B. MODEL-BASED HUMAN BEHAVIOR RECOGNITION APPLICATIONS

Currently, there are some model-based human behavior recognition applications using CSI. These applications utilize the physical law and mathematical model to depict the CSI variation and recognize human behavior. We make Table 7 interpret the main components of each application, including the device used, preprocessing method, test environment, number of users, classifier, and recognition performance. From this table, we discover that there exist many evident differences between model-based approach and pattern-based approach. Specifically, there are more daily activity recognition and user authentication applications using a pattern-based method while there are one or two

TABLE 7. Model-based human behavior recognition applications.

Type	Reference	Equipment	Sample	Preprocessing	Scenario	User	Model	Performance
Daily Activities Recognition	CARM [124]	Intel 5300 NIC	1400	PCA, Low-pass filter	hall, apartment room, office.	25	DTW, CSI-speed model, CSI-activity model.	The average accuracy is 96%.
	Zhang *et al.* [125]	Intel 5300 NIC	495	Savitzky-Golay filter	AP~PC: 5m.	11	Fresnel model, CNN.	Fresnel model: 95%, CNN: 82%.
Walking Direction Estimation	WiDir [129]	Intel 5300 NIC	856 + 433	Savitzky-Golay filter	6m × 7m, 6m × 6m, 4m × 3m.	5	temporal-spatial model, Fresnel zone model.	Median error: less than 10 degrees.
Human Presence Detection	FreeSense [126]	Intel 5300 NIC	291	———	Lab: 7m × 6m, bedroom: 5m × 7m, conference:7m×9m.	3	Wi-HD model, Fresnel zones model, AoA	FP: 0.53%, FN: 1.40%.
	RDFID [127]	Intel 5300 NIC	——	PCA, CWT	living room: 5m × 4m, office: 5m × 4m, 10m × 6m.	4	GMM	walk: FN: 2.3-FP: 1.7; bend: 2.5-1.7; creeping: 2.6-1.5
User Authentication	WiPIN [128]	Intel 5300 NIC	30 × 30	Butterworth filter	Lab 1, Lab 2, seminar room, meeting room, corridor.	30	Interacting Model, SVM	92%
Handwriting Recognition	WiDraw [130]	Atheros AR9590	8448	low pass filter	apartment, office, cafeteria.	10	AoA	Average 91%
Respiration Detection	Wang *et al.* [131]	Intel 5300 NIC	——	———	office, hall.	2	PSD, Fresnel model.	none
	TinySense [132]	Intel 5300 NIC	——	wavelet filter, mean filter, TOA filter.	home	4	Fresnel Zone model	1 user: 95%, 2 users: 88%.
	Yang *et al.* [134]	Intel 5300 NIC	——	Hampel and wavelet filter.	———	5	Gaussian model, and Fresnel zone model.	MAE: 0.5 bpm−1 bpm.
	Zhang *et al.* [135]	Intel 5300	——	FFT	AP-PC: 1m.	8	Fresnel diffraction model	98% (location more than 60% in FFZ).
	FullBreathe [24]	Rice WARP platform	10 × 5	CM	hall room, two bedrooms, living room.	8	Fresnel Zone and PSD.	Best: 100% (phase and amplitude of CM).
	BreathTrack [136]	Intel 5300 NIC	——	Hampel filter, FIR high pass filter	meeting room: 5m × 8m.	8	AoA	over 99% in most scenarios.
	FarSense [137]	Intel 5300 NIC	——	Savitzky-Golay filter	7.5m × 9m, AP-PC: 6~9m.	—	CSI-ratio model	17.9 bpm

similar applications using the model-based method. Conversely, the number of respiration applications based on the model far exceeds that of applications based on the pattern. The leading cause is the difference in the characteristics of these two methods. Specifically, pattern-based methods are suitable for recognition or inference of large range activity while model-based methods are fit to recognize the periodic activities with fine granularity. Besides, we notice that the types of recognized behavior are very different. Specifically, there are six types of model-based applications while there are about eighteen types of pattern-based applications. The latter outweighs the former. The reason may be that building an appropriate model is challenging because many factors affect the precision of the model. The typical model-based applications include daily behavior recognition [124], [125], human detection [129], walking direction estimation [126], and respiration detection [24], [131]–[136]. We interpret the key element of these applications as follows.

1) COARSE-GRAINED SPECIFIC BEHAVIOR RECOGNITION

DAILY ACTIVITY RECOGNITION

Wang *et al.* propose a CSI-based human activity recognition and monitoring system, called CARM [124] in 2015. Two powerful theory models are proposed: a CSI-speed model, which describes the relationship between CSI dynamic value and human movement speeds, and a CSI-activity model, which depicts the relationship between the movement velocity of human body parts and a special individual action. Furthermore, the authors propose a series of signal processing approaches, including PCA to de-noise, DWT to extract features that represent the movement speeds, and HMM to address different individual activity speeds. Furthermore, it uses the Baum-Welch algorithm to calculate the average vector and covariance matrix associated with each state and the transition probabilities of HMM. In addition, the authors apply the Exponential Moving Average (EMA) algorithm to adjust the detection threshold to determine the start or end of an activity. Based on these models and algorithms, the authors

deeply analyze the correlation between CSI value and human speed. The experiments with eight different actions in laboratory and apartment are used to validate the system. The results show that CARM can effectively detect small movements and large movements with a true positive rate (TPR) larger than 98% when the distances at 5 meters or 12 meters, respectively. Besides, it implements impressive mean activity identification accuracy of 96% across all activities in the trained environment and an identification accuracy for more than 80% in a new environment and with a new individual.

2) FINE-GRAINED SPECIFIC BEHAVIOR RECOGNITION

a: HANDWRITING RECOGNITION

Li *et al.* propose the first WiFi-based hand motion tracking system, called WiDraw [130] in 2015. This system can automatically track the hand's trajectory and estimate the letters written. The key structure of WiDraw consists of initial phase calibration and MUSIC algorithm. Specifically, the authors apply a laptop that equips Atheros 9590 to collect CSI data and then calibrate the phase of CSI stream. After that, the authors utilize a low pass filter to eliminate noise caused by environment inference. In addition, MUSIC's 1-D AoA model is adopted to track the hand's trajectory and realize the identification of drawn letters, words, and sentences. According to experimental results, WiDraw has fine performance of in-air handwriting (e.g., letters, words, and sentences) identification with an average 91% precision.

b: RESPIRATION DETECTION

Zhang *et al.* propose a novel contact-free breath tracking system in 2019, called BreathTrack [136]. This system can automatically track the status of respiration by exploiting Hampel filter, FIR high pass filter, and joint AoA-TOF sparse recovery method. Specifically, the authors apply the Hampel filter and FIR high pass filter (the cutoff frequency is 0.05Hz) to eliminate low-frequency noise. To avoid the phase distortions (e.g., Carrier Frequency Offset (CFO), Sampling Frequency Offset (SFO), Packet Detection Delay (PDD), and PLL Phase Offset (PPO)), the authors apply a combination of hardware and software methods. Furthermore, a joint AoA-TOF sparse recovery method is adopted to get the phase variation of attenuation coefficient and acquire respiratory state and breathing rate. The results show that BreathTrack achieves more than 99% breath detection precision in most scenarios.

3) ACTIVITY INFERENCE

a: WALKING DIRECTION ESTIMATION

Wu *et al.* propose the first CSI-based device-free human walking direction estimation system, called WiDir [129] in 2016. This system estimates a walking direction angle in real-time by analyzing the phase between two subcarriers. WiDir mainly consists of data acquisition, pre-processing, feature extraction, and direction calculation. Specifically, the authors apply the Savitzky-Golay filter to remove noise and smooth signals of CSI. After phase delay estimation and Fresnel direction estimation, a temporal-spatial model is applied to infer the walking direction. Extensive results demonstrate that WiDir can successfully detect the human walking direction with a median error of less than 10 degrees.

b: USER AUTHENTICATION

Wang *et al.* propose a novel device-free person identification system, called WiPIN [128] in 2018. This system can recognize user walking and infer the user's identity by utilizing the SVM and interacting model. The authors apply some techniques to process CSI data, including Butterworth filter to denoise and multipath effect removal based on IFFT-FFT frequency-domain method. After acquiring robust features that can represent walking, the authors utilize the interacting model and SVM to analyze the correlation between walking characteristics and the user's identity. According to extensive experimental results, WiPIN has 92% precision of the user's identity recognition over a group of 30 users.

c: HUMAN PRESENCE DETECTION

Xin *et al.* propose a novel device-free indoor human presence detection system, called FreeSense [126] in 2018. It not only implements human presence recognition but also models the sensing coverage for movements. This system can automatically detect indoor users by using the MUSIC algorithm, Fresnel zones model, AoA, and Wi-HD model, and have great anti-interference ability to resists noise, such as multi-path effect and device difference. Firstly, the authors apply the MUSIC algorithm to evaluate the phase difference between the waveforms in some receiving antennas. Furthermore, the Fresnel zones model, AoA, and Wi-HD model are utilized to estimate the coverage size and the sensing granularity. The experiments demonstrate that FreeSense achieves an average 0.53% false positive rate (FP) and an average 1.40% false-negative rate (FN); moreover, the coverage estimation approach of FreeSense has an average accuracy of 1.36 m.

C. DEEP LEARNING-BASED BEHAVIOR RECOGNITION APPLICATIONS

Currently, there have some deep learning-based human behavior recognition applications using CSI. These applications leverage the striking advantages of deep learning to recognize human behaviors. We make Table 8 to analyze the key features of recognition applications using deep learning from the following aspects, including the device used, preprocessing method, test environment, number of users, classifier, and recognition performance. From this table, we notice that the number of recognized behavior types using deep learning is between that of model and pattern. We deem that deep learning is becoming an important research method in behavior recognition using CSI. From recognition approaches, we find that most applications apply common deep models (e.g., Autoencoder, CNN, LSTM, RNN, and ResNet), which indicates that the general deep learning algorithm can be used at CSI-based behavior recognition and the

TABLE 8. Deep learning-based human behavior recognition applications.

Type	Reference	Equipment	Sample	Preprocessing	Scenario	User	Classifier	Performance
Daily behavior recognition	WiSDAR [138]	Intel 5300 NIC	Train：5760	low pass filter, PCA, STFT.	Lab: 8.2×5.4m^2, hall: 40 × 10m^2, Apart: 35m^2, office: 18m^2.	6	CNN, LSTM	96%
	DeepHare [139]	Atheros	Train : 8000 × 5× 10	low-pass or median filter.	conference: 5m × 7m, office: 5.6m × 9m, Apart:7.5m×8m	10	AE-LRCN	97.6%
	DeepSense [140]	Atheros	Train : 8000 × 5× 10	low-pass or median filter.	conference room: 5×7m^2, Apart: 7.5×8m^2	10	AE-LRCN	97.4%
	WiSPPN [142]	Camera, WiFi	Train 79496 + Test 19931	————	9.4m × 6.15m; 9.4m × 14.3m	8	FCN	PCK@50: 0.9059
Falling Behavior Detection	DeepFalls [143]	Intel 5300 NIC	1260	interpolated, HHT, CEEMDAN.	meeting room: 5m × 8m, Apart: 5× 12m^2	——	DCNN	7.7% and 11.9% higher than RT-Fall.
	WmFall [144]	Intel 5300 NIC	1000	Butterworth low-pass filter	Lab: 6m × 8m	6	SVM, DNN	89% precision with 8% false alarm
	ToiFall [145]	Intel 5300 NIC	1750	Gabor filter	6m × 6m	17	ResNet, LIBSVM	98.6%; 99.1%
Hand Gesture Recognition	DeNum [146]	Intel 5300 NIC	————	Band-pass filter, phase calibrate.	office	5	SVM, 4-layer DNN	94% in office
	DeepNum [147]	Intel 5300 NIC	6000	HOSVD	meeting:6×6m^2, office:6×12m^2, corridor:2×20	10	7-layer CNN	0 to 9 finger gestures: 98%.
	Temporal Unet [149]	WiFi, Camera	Train:1116 + Test: 278	————	indoor	1	DNN, DTW, KNN	AP@0.5: detection:1, classification:0.99
	FreeGesture [8]	Atheros	120 × 5	————	conference room	1	CNN	95.8%
Sign Language Recognition	SignFi [150]	Intel 5300 NIC	8280	Removing phase shifts	Lab: 13m × 12m, home: 4.11m × 3.86m.	5	CNN, KNN, DTW	Lab :98.01%, home: 98.91%, Lab + home 94.81%.
Gait Sensing and Walking	Xu *et al.* [151]	Intel 5300 NIC	Train: 67452 × 0.8; Test: 67452 × 0.2	IFFT, PCA, Butterworth bandpass filter.	Lab: 9.5m × 7.8m, area: 6.0m × 6m.	11	RNN	8 walking directions: 96.92%.
Direction Estimation	Jakkala *et al.* [152]	Intel 5300 NIC	1303	PCA, window function.	experiment room	30	DCNN	97.12±1.13%
Human Presence Detection	Wang *et al.* [153]	WiFi, camera	Train: 123631 + Test: 30996	————	class room, Lab, office.	8	Mask R-CNN, OpenPose	AP@50: 0.91
	CSI-Net [154]	Intel 5300 NIC	biometrics:43 077, detect: 43077, sign: 23896, fall: 24398	median filter, mean filter, Butterworth filter, DWT	one room with 3 kinds of arrangements.	30	DNNs, LibSVM, Naïve Bayes	Best: human identification: 93%, sign recognition: 100%, fall detection: 96.67%.
	Fang *et al.* [155]	Intel 5300 NIC	Train: 8000	calibration	bedroom:12m^2, 14m × 10m	——	DNN	93%
Crowd Counting	Chen *et al.* [156]	Intel 5300 NIC	train 10937 + test: A-10565 +B-11818	————	conference: 5.8m × 8.6m.	9	DNN	88%
	WiCount [157]	Intel 5300 NIC	Total:158333	Butterworth filter, PCA.	Lab 1, 2	10	DNN, SVM	5 users: 82.3%, 10 users: 75%.
	Doong *et al.* [158]	Intel 5300 NIC	train 12800 + test 3200	autoencoder	experiment room	5	SVM, RF, DNN	DNN :0.827, SVM: 0.78, RF: 0.739.
	DeepCount [159]	Intel 5300 NIC	800	phase sanitization	Lab1,2	10	CNN-LSTM	86.4%, 90%
User Authentication	Shi *et al.* [160]	Intel 5300 NIC	3336 + 834	Butterworth low-pass filter	office: 26ft × 14ft, apartment: 36ft × 22ft.	11	SVM, DNN	Accuracy: over 94% (walking) and 91% (stationary activities)
	WiAU [161]	Intel 5300 NIC	0.6, 0.2, 0.2	Butterworth low-pass filter	office, corridors.	12	ResNet, CNN	98%
	NeuralWave [162]	Intel 5300 NIC	0.7, 0.3	wavelet transform, PCA	Lab	24	ConvNet, RadioNet	87.76 ±2.14%
	FingerPass [163]	Intel 5300 NIC	————	Butterworth filter	living room: 5.8m × 4.2m, bedroom: 3.8m × 3.4m,kitchen: 3.4m × 2.2m	7	LSTM SVDD	94.1%
Respiration Monitoring	Khan *et al.* [22]	USRP B200	1600	adaptive filter	Static and random motion scenario.	3	random forest, CNN	CNN: 94.85%, 81.03%,85.17; RF: 1,71.07%,81.97.

capability of automatic feature extraction and classification can be exploited when developing these systems. We believe that more and more CSI-based behavior recognition applications will employ deep learning algorithms because it can build and extract the high-level behavior features that usually cannot be extracted by pattern-based or model-based methods. The typical deep learning-based applications include daily behavior recognition [138], [139], [141], [142], falling detection [143], [144], syncope detection [145], hand gesture recognition [8], [146], [147], [149], sign language recognition [150], gait and walking direction recognition [151], [152], human detection [153], [154], crowd counting [156]–[158], user authentication [161], and respiration monitoring [22]. We explain the key components of these applications as follows.

1) COARSE-GRAINED SPECIFIC BEHAVIOR RECOGNITION

a: DAILY BEHAVIOR RECOGNITION

Wang *et al.* propose a novel device-free spatial diversity-aware activity recognition system, called WiSDAR [138] in 2018. This system identifies eight kinds of activities (e.g., walking, falling, running, sitting, picking, pushing, waving, boxing) in four environments (e.g., laboratory, hall, apartment, office) by using Low-pass filter, PCA, STFT, CNN, and LSTM. Specifically, because daily activities cause more low-frequency components of CSI streams, the authors apply low pass filter and PCA to wipe out high-frequency noise. Moreover, the STFT is adopted for feature extraction. Finally, the authors utilize CNN and LSTM to train 5760 CSI samples collected from six students. The sufficient experimental results show that WiSDAR achieves 96% stable identification precision for eight kinds of activity identification. Zou *et al.* propose a novel device-free human behavior recognition system, called DeepHare [132] in 2018. This system can automatically identify 5 daily behaviors (e.g., sit, stand, walk, run and lying down) in three different rooms (e.g., conference room, office, apartment) by utilizing Autoencoder Long-term Recurrent Convolutional Network (AE-LRCN). Specifically, the authors apply the Atheros CSI Tool and OpenWrt to acquire CSI measurements, and then the low-pass filter and median filter are applied to eliminate inherent noise. Besides, the authors utilize AE-LRCN to train 400 thousand CSI samples and realize activity recognition. According to experimental results, DeepHare can identify 5 types of daily actions with 97.6% accuracy.

b: FALLING DETECTION

Chowdhury *et al.* propose DeepFalls [143] in 2017, a fall detection based on Wi-Fi spectrograms and deep convolution nets. This system first linearly interpolates the signal and then used Hampel identifier to wipe out outliers of CSI stream. Afterward, de-trending subcarriers, zero-padding, and tapering CSI waveforms are applied as the signal segmentation techniques for the detection of fall and fall-like activities. Next, this system applies the Singular Spectral Analysis (SSA) to remove the noise of CSI amplitudes rather than using DWT. Due to the shortcomings of STFT and Continuous Wavelet Transform (CWT) spectrograms, the authors decompose the input signal to Intrinsic Mode Functions (IMFs) with the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN). Then, the Hilbert Huang Transform (HHT) of the decomposed IMFs is applied to create spectrogram images, which are fed into 10-layer Deep Convolutional Neural Networks (DCNNs) for classification. The extensive experiments demonstrate that DeepFalls achieves satisfactory recognition performance compared with RT-Falls. In all environments, DeepFalls achieves a higher sensitivity (7.7%) and specificity (11.9%) on average than RT-Fall. When the furniture position is changed in the same room, DeepFalls implements a 4.07% higher sensitivity and 14.66% higher specificity than RT-Falls. When training on one scenario and testing in different scenarios, the sensitivity and specificity of DeepFalls are much lower than before. In this scenario, DeepFalls has 9.67% and 12.3% improvement in sensitivity and specificity than RT-Falls.

2) FINE-GRAINED SPECIFIC BEHAVIOR RECOGNITION

a: HAND GESTURE RECOGNITION

Zhou *et al.* propose a novel device-free real-time finger gesture recognition system in 2018, called DeepNum [147]. This system can automatically detect ten kinds of number finger gestures in three environments (e.g., meeting room, corridor, student studio) by using Higher-Order Singular Value Decomposition (HOSVD) and CNN. Specifically, the authors apply HOSVD to wipe out the noise of the CSI stream and acquire useful high-dimensional principal components. In addition, the authors utilize a 7-layer CNN to train 3000 ($300 \times 20 \times 50\%$) samples of CSI and realize finger gesture recognition. According to the experimental results, DeepNum has 98% accuracy for 10 finger gesture identification.

b: SIGN LANGUAGE RECOGNITION

Ma *et al.* propose a sign language recognition system, called SignFi [150] in 2018. This system can identify more than 25 hand gestures by using 9-layer CNN. Specifically, because CSI phase change in the range of $-\pi$, π, phase calibration becomes necessary. Besides, a 9-layer CNN and KNN with DTW are utilized to identify 276 sign gestures. According to the experimental results, when SignFi validates system performance with one user in three places, such as laboratory, home, and both laboratory and home environments, this system has great performance with the accuracy of 98.01%, 98.91%, and 94.81%, respectively. In addition, SignFi achieves 86.66% mean precision for the recognition of 150 sign gestures performed by five different volunteers.

c: RESPIRATION DETECTION

Khan *et al.* [22] propose a novel device-free end-to-end deep learning-based approach to detect human respiration rate using CSI in 2017. The main components of this approach consist of adaptive cancellation, deep activity classification,

and breathing rate estimation. Specifically, the authors apply USRP B200 SDR to collect the CSI data stream and then utilize an adaptive filter to wipe out the redundant echoes of signals. Furthermore, the authors utilize Random Forest (RF) and CNN to monitor the human breathing rate in three different environments, such as breathing motion environment, static environment, and random motion environment. In a static environment, the breathing rate detection performance with RF is pretty great than that with CNN, and RF has 100% accuracy whereas CNN achieves 94.85% accuracy. In other environments, the CNN model has higher robustness than RF and the performance of CNN outweighs that of RF. Specifically, CNN achieves the recognition accuracy of 81.03% and 85.17% while RF obtains 71.07% and 81.97% accuracy in random motion and breathing motion environment, respectively.

3) ACTIVITY INFERENCE

a: GAIT SENSING AND WALKING DIRECTION ESTIMATION

Xu *et al.* propose a novel device-free cycle-independent human gait and walking direction detection system in 2018 [151]. This system can recognize human gait and walking direction by using Inverse Fast Fourier Transformation (IFFT), Butterworth bandpass filter, PCA, STFT, and attention-based RNN encoder-decoder neural network. Specifically, it requires three steps to handle raw CSI, including long delay removal using IFFT, CSI de-noising using Butterworth bandpass filter, and CSI refining using PCA. In addition, the authors apply the attention-based RNN encoder-decoder neural network to determine human gait and walking direction. The results show that this system can recognize human gait from a group of eight volunteers with 89.69% average F1 scores and identity direction from eight walking directions with 95.06% accuracy. Moreover, the mean recognition accuracy considering both direction and human gait exceeds 97%.

b: HUMAN PRESENCE DETECTION

Wang *et al.* propose an innovative deep learning-based activity recognition system in 2019, called CSI-Net [154]. This system comprises four aspects: biometrics estimation (including body fat, muscle, water, and bone rates), human identification, hand sign recognition from 0 to 9, and falling detection. Firstly, the authors adopt mini-PC which equipped with Intel 5300 NIC to collect a great number of samples of CSI (e.g., biometrics estimation: 43077, human identification: 43077, hand sign recognition: 23896, and falling detection: 24398). Next, the median filter, mean filter, Butterworth filter, and DWT are applied to wipe out the noise, and then the Local Temporal Average is employed to smooth CSI streams. Finally, the authors apply DNNs, LibSVM, and Naïve Bayes to classify human activities. The result shows that DNNs has better identification performance (e.g., human identification: 93%, sign recognition: 100% and fall detection: 96.67%.) than LibSVM (e.g., human identification: 85.28%, sign recognition: 90.24 % and fall detection: 81.46%.), and Naïve Bayes (e.g., human identification: 72.97%, sign recognition: 81%, and fall detection: 73.01%.).

c: CROWD COUNTING

Liu *et al.* propose WiCount [157] in 2017, a robust deep learning-based crowd counting system by identifying walking and other specific activities (e.g., eating, waving, typing, talking, sitting down). This system uses Butterworth filter, weighted moving average filter, PCA to wipe out high-frequency noise of CSI stream. Next, SVM and Back Propagation Neural Network (BPNN) are applied to realize crowd estimation. According to experimental results, WiCount can automatically estimate the number of crowds up to five and ten with 82.3% and 75% precision, respectively. In addition, the experiments show that the combination of amplitude and phase information as behavior features can improve crowd counting precise and deep learning algorithm performs better than SVM for crowd counting.

d: USER AUTHENTICATION

Lin *et al.* propose a device-free user authentication system, called WiAU [161] in 2018. This system can automatically recognize 16 kinds of activities (e.g., walk, one arm wave, high arm wave, sit down, drink water, squat, phone call, etc.) and infer one user's identity according to walking gait. Specifically, the authors apply Butterworth low-pass filter to remove the high-frequency noise of raw CSI stream. Furthermore, Automatic Segmentation Algorithm (ASA) is used to segment continuous behaviors. Finally, the authors apply a convolution module that contains one CNN layer and 15 ResNet layers to prove user authentication (e.g., legal users and illegal ones). In this experiment, the authors invite 12 volunteers (9 males and 3 females) to participate in experimental data collection. The data ratio of training, validation, and testing is 60%, 20%, and 20%, respectively. Experimental results confirm that WiAU achieves over 98% precision for identity identification.

V. DISCUSSION

In this section, we make a detailed discussion on certain kinds of behaviors from different aspects, including test environment, size of samples, and system performance. This discussion section focuses on analyzing the characteristics of these behaviors and comparing the difference of the system implementation. We hope this part provides some insights into the development of human behavior recognition using CSI. Based on the representative applications, we analyze some main behaviors (see in Fig. 12), such as daily behavior recognition, falling detection, hand gesture identification, crowd counting, and user authentication. Since almost all systems employ Intel 5300 NIC as CSI collection device, we do not emphasize the hardware device in the following discussion.

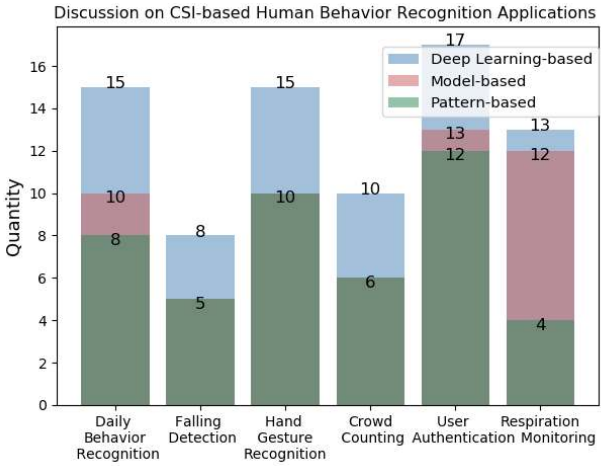

FIGURE 12. Recognition technique statistics of the different human behaviors from Table 3.

A. DISCUSSION ON DAILY BEHAVIOR RECOGNITION

Daily behavior recognition is an important research field because it can be utilized to identify human life state and provides helpful information to evaluate lifestyle. Daily behaviors have been widely studied and can be recognized using different recognition methods. The main reasons may be explained as follows. Firstly, the range of daily behavior is bigger than that of other behaviors. Therefore, the variation of CSI is easily measured and the different behaviors are easily discriminated. Secondly, daily behavior recognition can be treated as a classification problem that can be effectively tackled. Therefore, we find that many applications of daily behavior recognition employ the pattern-based method rather than model-based methods.

We analyze the main characteristics of daily behavior recognition and make some comparisons between recognition methods. Firstly, from the size of samples, the number of samples of the pattern-based methods is less than 5000 and the median value is 2000. The samples of model-based methods are usually 1500 which is less than that of the pattern-based methods. Furthermore, the samples of deep learning-based are above 5500 except Khan *et al*. [141] and the maximum value reaches one hundred thousand. Secondly, we observe that the activities are conducted usually in 2 to 4 scenarios to assess system performance and the type of activities is usually 5 to 9, which validates the robustness of the system. The overall performance of daily behavior recognition is close to 95% for three recognition techniques, which indicates that different recognition methods are not a decisive factor. Although the performance of deep learning-based is a little higher than other methods, it usually requires more computation and more data to train neural network parameters, which is the main disadvantage. The best recognition results are shown as follows. DeepHare[139] has 97.6% precision using the deep learning-based approaches, WiHACS [57] achieves 97% precision using the pattern-based methods, and CARM [124] applies model-based methods and has 96% accuracy. Thirdly, in through-the-wall scenarios, the identification accuracy may decrease, such as WiHACS [57], DFS [60], and TW-See [61]. For a special test environment, Wei *et al.* [56] consider the effect of radio-frequency interference that seriously decreases recognition precision. Besides, WiSPPN [142] studies human behaviors by analyzing human pose and calculates key point coordinates instead of behavior types; therefore, it applies a different evaluation metric to assess the system performance.

B. DISCUSSION ON FALLING DETECTION

Falling is a serious threat to the life of persons, especially for the elderly. Therefore, many researchers pay more attention to the falling detection using CSI due to its many advantages. Currently, there some studies which can automatically detect falling behaviors, such as WiFall [67], Anti-Fall [68], RT-Fall [69], FallDeFi [70], Dong *et al.* [71], DeepFalls [143], and WmFall [144]. Based on the analysis of these applications, we obtain the following results. Firstly, for the falling detection, the pattern-based method is more popular than the deep learning-based method. To validate system performance, many applications conduct the behaviors at 2-3 test environments. In addition, we notice that some applications of falling detection consider falling as one of the recognized behaviors while some applications solely identify falling activity, such as DeepFalls and FallDeFi. Secondly, we discover that the number of samples of falling detection is similar to that of daily behavior when using pattern-based methods because some falling detection applications consider falling as a common behavior. The samples of falling have less size compared with that of daily behavior using deep learning because these systems usually consider whether the behavior is falling or not, which is a simple binary classification problem. Thirdly, the recognition accuracy of falling is lower than that of daily behavior using the pattern-based and deep learning-based approach because the duration of falling is very short, which leads to the difficulty of the detection and identification. Therefore, many applications add another evaluation parameter, false alarm, to provide a more precise assessment. The recognition accuracy of pattern-based and deep learning-based hold similar results, about detection accuracy of 90% and the false alarm rate of 12%. Some typical applications make comparisons with other applications to confirm system improvements. For instance, WiFall achieves 87% detection precision with a false alarm rate of 18%. RT-Fall compares its performance with WiFall and has a 14% higher sensitivity and 10% higher specificity than WiFall. Furthermore, FallDeFi and DeepFalls compare their performance with RT-Fall. FallDeFi and DeepFalls are more sensitive to detect falling than RT-Fall. FallDeFi achieves 93% accuracy of falling detection, compared with RT-Fall and CARM, this system improves 12 % and 15 % accuracy, respectively. DeepFalls has an average 7.7% higher sensitivity and 11.9% higher specificity than RT-Fall. Dong *et al.* [71] realize falling detection in the staircase with 94% precision.

The results indicate that the test environments affect the recognition accuracy of falling detection.

C. DISCUSSION ON HAND GESTURE RECOGNITION

Hand gesture recognition is an important field of human behavior recognition because it can provide effective information input for HCI (Human-Computer Interaction) and more help for communication of the deaf. Currently, there has a lot of systems which identify various hand gestures, such as WiG [73], WIGeR [74], Mudra [75], WiFinger [76], WiFinger [77], PWiG [78], Chen *et al.* [79], WiCatch [80], iGest [81], DeNum [146], FreeGesture [8], DeepNum [147], Widar3.0 [148], and Temporal Unet [149]. From the recognition technique, we find that the pattern-based method is more popular than the deep learning-based method for hand gesture recognition. The experimental scenarios of these applications are simple compared with daily behavior. From the size of samples, the number of samples of the pattern-based is less compared with the deep learning-based. The size of samples used in pattern-based approaches is less than 500 (except WiG [73]) while the number of samples used in deep learning-based approaches is over 1000 for common hand gestures (e.g., DeepNum [147]: 6000, Widar 3.0 [148]: 1700). From system performance, the overall recognition accuracy is about 93% for LOS scenarios and less 90% for NLOS scenarios, such as WiG [73] and PWiG [78]. DeepNum has an overall accuracy of 98% for American sign language, which is more accurate than other sign language applications. WiCatch [80] considers two hand gesture recognition and achieves 95% recognition accuracy. We notice that the number of hand gestures varies from 4 (e.g., WiG [73]) to 9 (e.g., Mudra [75]) and the size of ASL (digital gestures) is up to 10.

D. DISCUSSION ON CROWD COUNTING

Crowd counting is a significant study because it can provide helpful information on population mobility and human dynamics. Based on the information, we can conduct some intelligent crowd management when person density reaches some thresholds. Currently, there has many applications which estimates the number of users according to different walking states, including FCC [98], FreeCount [99], Guo *et al.* [100], WiFree [101], Wi-Count [102], HFD [103], Chen *et al.* [156], Wi-Count [157], Doong *et al.* [158], and DeepCount [159]. The pattern-based systems have more applications than deep learning-based systems. We observe that the size of samples of pattern-based is less than 1000 while that of deep learning-based is above 16000. The experimental scenarios are very different, varying from 4 different environments (FCC [98]) to a single environment (Doong [158]). The recognized number of crowds varies from 5 users (WiCount [157]) to 15 users (FCC [98]). All these applications recognize walking users and the recognition precision decreases with the increase of users. From the perspective of recognition performance, the pattern-based methods have better results, and the counting accuracy is higher than 90% while all deep learning methods are less than 90%, which indicates pattern methods are suitable to crowd counting applications.

E. DISCUSSION ON USER AUTHENTICATION

User identity authentication is an important premise for almost all applications because it can prevent illegal access to private information. Currently, a great deal of applications based on CSI realize user's identity recognition, including WiFi-ID [112], WiWho [113], WifiU [114], FreeSense [115], Wii [116], Chen *et al.* [117], AutoID [118], Wi-Sign [119], Liu *et al.* [120], BodyPIN [121], BioID [122], WiPIN [128], Shi *et al.* [160], and WiAU [161]. Many user authentication systems apply pattern-based methods except WiAU [161], NeuralWave [162], FingerPass [163], Kong *et al.* [163], and WiPIN [128]. The authentication number of users varies from 2 (e.g., WiFi-ID [112]) to 30 (e.g., BodyPIN [121], WiPIN [128]). The size of samples of these applications is similar to the falling detection applications. Most experimental environments contain 2-3 kinds of scenarios, such as the office, laboratory, and apartment, etc. Like crowd counting, most authentication applications utilize walking to determine user identity except BioID [122] and Wi-Sign [119], which utilities lip-motion and hand-motion to identify user's identity, respectively. The best recognition accuracy is Wii [116], which achieves 98.7% for two user's authentications. Other remarkable systems are BodyPIN [121], and WiPIN [128] because these systems can determine 30 users with 92% recognition accuracy. WiAU [161], applying a deep learning method, achieves 98% recognition accuracy for 12 users, which is a satisfactory result for crowd counting.

F. DISCUSSION ON RESPIRATION MONITORING

Normal respiration plays a crucial part in daily life because abnormal respiration state may endanger a person's life. Currently, plenty of systems based on CSI can detect human breath rate, including PhaseBeat [85], TR-BREATH [87], Liu *et al.* [88], Wang *et al.* [131], TinySense [132], Yang *et al.* [134], Zhang *et al.* [135], FullBreathe [24], BreathTrack [136], FarSense [137], and Khan *et al.* [22]. Different from other behavior recognition that usually applies the pattern-based or deep learning-based methods, most respiration monitoring applications leverage model-based methods. The number of users varies from 2 (e.g., TinySense [132]) to 12 (e.g., TR-BREATH [87]). The pattern-based method achieves the best results for the number of users (12 users) and recognition accuracy (above 98% for doze users under LOS and 9 users under NLOS scenarios). BreathTrack [136] monitors 8 users and achieves over 99% accuracy in most scenarios. For deep learning-based methods, Khan *et al.* [22] leverage CNN and obtain 98.85% recognition accuracy for 3 users, which is a common result because it collects more measurement data using USRP B200.

All in all, the pattern-based method is widely employed in most behavior recognition applications since it has many important advantages, such as fewer sample requirements compared with the deep learning-based approach, wide and

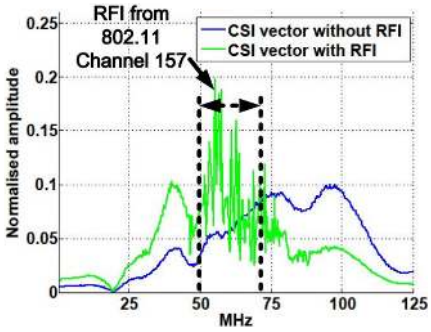

FIGURE 13. CSI without and with RFI [56].

deep study, simple recognition steps, and without the need of the precise mathematical model. Model-based approaches are extensively exploited in respiration monitoring due to the small range of body motion. In other words, the physical model is more appropriate for the detection of minor body motion. Deep learning is gradually being applied in CSI-based human behavior recognition because it has achieved extraordinary success in various scenarios, including image processing, speech recognition, and natural language processing, etc. Thereby, we hope that these CSI applications can leverage its powerful feature extraction and recognition capability to identify human behaviors. Generally, deep learning approaches need a large number of samples to extract features and implement classification, which is not absolute requirements based on our analysis. Thereby, if we have more samples, we can leverage the strengths of deep learning. If we have small size samples, we can select appropriate recognition algorithms based on characteristics of the system. Therefore, we deem that the size of samples affects the selection of recognition algorithms. In summary, we should select appropriate recognition methods to identify human behavior based on application requirements, experimental environments, and characteristics of algorithms.

VI. ISSUES AND FUTURE RESEARCH DIRECTIONS

Recently, human behavior recognition based on CSI has achieved remarkable success in many fields due to the popularity of WiFi devices and the improvement of recognition algorithms. However, we have to face many challenges when developing specific behavior recognition applications. Besides, with the gradual changes in requirements and work environment, we have to tackle more complex problems. In this section, we discuss some crucial issues and present some promising research directions. We hope that these contents provide some insights into the analysis of activity identification systems and facilitate the development of novel applications. We consider some representative problems and research trends, including electromagnetic interference, multiple users, through the wall, multiple AP, standard dataset, robustness, and security issues. These issues and further directions are discussed as follows.

A. ELECTROMAGNETIC INTERFERENCE

Although the dense deployment of WiFi brings us a more convenient network link, it causes serious electromagnetic interference, which reduces the accuracy of behavior recognition. As shown in Fig. 13, Wei *et al.* [56] exhibit a device-free activity recognition system with radio frequency interference (RFI), which can identify 4 kinds of behaviors (e.g., lying, sitting, standing, and walking). Based on this system, authors find that the CSI signal has been seriously affected and the measurement data have been apparently changed by RFI. Consequently, the behavior recognition accuracy decreases due to electromagnetic interference. Huang *et al.*consider the co-channel interference from channel overlap of WiFi devices and propose WiAnti [196], a robust anti-interference activity recognition system using CSI. WiAnti analyzes the co-channel interference and proposes a subcarrier selection algorithm to choose some subcarrier with a weak correlation. The system achieves 95.865% recognition accuracy, an 8% improvement compared with WiFall. With the increase of WiFi devices, how to reduce electromagnetic interference and improve the recognition accuracy may not be a nonnegligible issue. The effect of RFI on system performance should be a factor considered when developing and evaluating a gesture identification based on CSI.

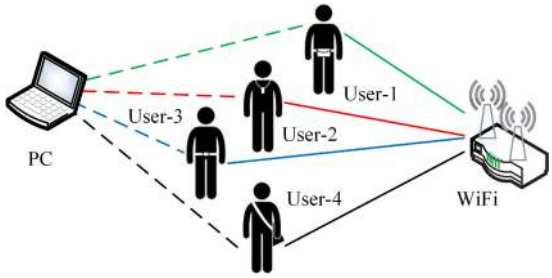

FIGURE 14. Multi-person behavior recognition.

B. MULTI-PERSON BEHAVIOR RECOGNITION

Multi-person behavior recognition refers to that the system can recognize the gestures conducted by more than one person simultaneously, as shown in Fig. 14. Specifically, when multiple persons perform actions in the coverage of CSI signal, besides the effect of gesture conducted by one user on the CSI, the mutual position of users also pose more complicated influence on CSI signal, which extraordinarily increases the difficulties of recognition because the user may walk to different positions. However, multi-person behavior identification is also indispensable for some applications. Thereby, some applications evaluate system performance by analyzing the test scenario. For instance, WiMU [59] is a multiple user gesture recognition system using a WiFi signal. It utilizes WiFi signal propagation law to build a theoretical model for depicting multiple user movement. It achieves average accuracies of 95%, 94.6%, 93.6%, 92.6%, and 90.9%

for 2, 3, 4, 5, and 6 simultaneously performed gestures, respectively. Tan *et al.*propose MultiTrack [62], a multi-user tracking and activity recognition system. It extracts signal reflection describing each user using multiple WiFi links and achieves over 92% precision for activity recognition under multi-user scenarios. WiHear [89] identities 9 vowel pronunciation according to mouth motion profile, and it has 91% precision for one person and 74% precision for 3 persons. TinySense [132] detects a multi-person breathing rate by analyzing biggest stream with the peak-valley difference. Moreover, this system obtains over 95% and 88% precision for respiration rate detection of 1 or 2 volunteers at the same time. Yang *et al.* [134] propose a multi-person sleeping respiration monitoring system. This system estimates two people's breathing rate with the Mean Absolute Error (MAE) of 0.5bpm∼1bpm. TR-BREATH [197] detects the respiratory rate of 1 to 7 people. In this system, the accuracy of a single breathing estimation is 98.5%, and the average accuracy of 1 to 7 respiratory estimation is 96.9%. We discover that the increase in the number of users usually decreases the recognition accuracy. How to effectively address the effect of multiple persons and recognize different actions conducted by different users is a challenging problem.

C. THROUGH THE WALL DETECTION

CSI signal has a remarkable benefit because it can propagate through the wall. The communication between the sender and the receiver that locate in different rooms do not be interrupted if the distance between them keep a rational range. Consequently, human behavior recognition in a through-the-wall scenario by using CSI extends the research field and provides us with promising applications. With the powerful communication through the wall, many systems have implemented attractive functions, such as E-Eyes [55], TensorBeat [86], WiHACS [57], WiseFi [176], Smokey [97], NotiFi [66], R-TTWD [108], WIGeR [74], FallDeFi [70], DFS [60], and TW-See [61]. However, from these through-the-wall activity recognition applications, we find that only DFS [60] and TW-See [61] involve the material of walls, such as concrete walls and glass walls. For instance, DFS [60] identities 8 kinds of actions in two experiments with exceeds 85% precision. As shown in Fig. 15 (a) and Fig. 15 (b), two-sided walls, such as the concrete wall, and the glass wall and concrete wall, isolates receiving end and transmitting end. TW-See [61] recognizes seven daily behaviors (e.g., walking, hand swing, boxing, etc.) with an average 94.46% recognition precision in two through-the-wall scenarios (e.g., through the glass wall, and through the concrete wall), as shown in Fig. 16. These systems prove that the material of walls can affect signal propagation and lead to different recognition accuracy. Although a few applications discuss the recognition performance in through-the-wall scenarios, they do not provide a comprehensive analysis of variation of through-the-wall CSI signal. The applications and analysis of human behavior recognition in the through-the-wall scenario will be the hot research topic.

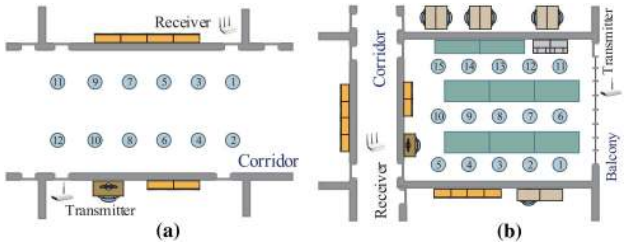

FIGURE 15. "Through wall" experiments of DFS [60]. (a) Corridor with two concrete walls. (b) Experiment with one glass wall.

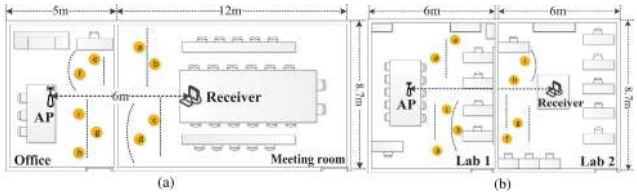

FIGURE 16. "Through wall" experiments of TW-See [61]. (a) Experiment 1 with one concrete wall. (b) Experiment 2 with one glass wall.

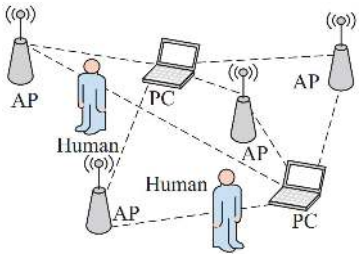

FIGURE 17. Wi-Fi-based activity recognition with multiple APs [198].

D. MULTIPLE ACCESS POINTS (APS)

Due to the short WiFi communication distance and low transmission power, the traditional CSI-based behavior recognition applications often use a single AP to validate algorithm accuracy [198]. However, various factors (e.g., wall, noise, etc.) severely distort the signal propagation and attenuate signal energy, which may decrease recognition accuracy. With the popularity of WiFi devices, multiple access points are available in our daily environments. Therefore, utilizing multiple access points seems to be a potential solution to improve recognition accuracy. As shown in Fig. 17, data collection with multiple APs provides us with more information from more communication links, which enhances behavior recognition accuracy in complex experiment environments. For instance, NotiFi [66] adopts five APs to evaluate the impact of the number of AP on behavior recognition accuracy. The experiments in NotiFi confirm that the increase in the number of WiFi AP can improve system performance. However, the relationships between recognition accuracy and the number of APs remain unclear. Li *et al.* [198] propose a learning method that can analyze the CSI of multiple APs. This approach adopts 9 APs to collect data of CSI and then utilizes CNN to identify human activities. It proves that using multiple APs for human activity recognition can increase the

identification precision effectively. However, how to decrease the RF interference from multiple APs, how to assign the position of the APs to obtain the optimal recognition accuracy, and how to coordinate the communication among these APs are the crucial problems to be solved.

E. STANDARD DATASET

Currently, almost all behavior recognition applications evaluate system performance using specific samples. Usually, authors recruit some participants to conduct some typical actions and collect the CSI data. The experiential environments are settled according to the specific requirements of the application. Consequently, system performance usually is validated under their own system arrangement. The comprehensive evaluation and comparison among more systems are difficult because the experiment conditions are very different. Although some applications provide some comparison with other applications, more analysis and discussion are needed. The standard dataset can significantly improve system performance evaluation and comparison. With an open and accurate dataset, we can accurately assess the system performance from many different aspects. As for how to build the dataset, we can obtain some experience from other open and successful datasets, such as CIFAR10 [199], ImageNet [200], MNIST [201]. When building the dataset, many factors should be considered, such as action types, number of participants, the difference of users, test environments, and size of samples. We deem that the successful development of the standard dataset will boost the studies and applications of behavior recognition based on CSI.

F. ROBUSTNESS

Many behaviors of current applications are conducted in indoor environments. As a result, the environment exerts an important effect on recognition accuracy [183]. Besides, the number of participants, age of the users, the position of devices, and types of gestures will affect recognition accuracy. How to make the algorithm available under different indoor scenarios and for different users is an essential issue because it confirms whether the approach has robustness. Some applications seek the solution using a model-based or deep learning-based approach. They can alleviate the problem and improve recognition accuracy under different environments. However, with these two approaches, we have to face the difficulty of modeling or gathering large samples. Therefore, how to tackle this question has no clear answer. Currently, we can take some measures to mitigate the effect of different test environments. Development of a universal framework to tackle this problem seems to be a potential approach [8], [22], [138], [141], [147], [150], [161], [187], [188]. Keeping robust and available for many scenarios is a fundamental problem for behavior recognition using CSI and can be solved in further research.

G. SECURITY ISSUES

The human behavior recognition based on CSI works under device-free and non-intrusive pattern, which provide a long-term and accurate monitoring of a user. The behavior recognition brings us many advantages and disadvantages. On the one hand, it provides us with more control over the smart device, timely health care to the elderly, and more help for the impaired person, which are their advantages. On the other hand, it may bring many disadvantages [202]. For instance, it can be utilized to steal private information by recognizing keystroke information [91], inferring a text-based password [92]. Besides, continuous behavior recognition actually poses strict surveillance on the user, which may lead to a serious threat to people's privacy if this information is utilized illegally. Furthermore, it can be leveraged to control the device remotely without permission in a quiet state. Therefore, how to control device only under the authority or how to prevent private information leak using CSI are essential problems.

VII. CONCLUSION

Human behavior recognition technology is an important research direction in the field of ubiquitous computing. Currently, human behavior recognition based on WiFi CSI has drawn more attention because it can overcome the shortcomings of traditional methods, such as the requirement of wearing physical sensors, privacy violations, and deployment costs. As a result, much significant research progress has been achieved in many application fields. This paper investigates state-of-the-art behavior recognition applications based on CSI and presents a comprehensive review of the key characteristics of these applications.

Firstly, this paper introduces the current general identification methods of behavior recognition, overviews related surveys, introduces the concept of channel state information, and illustrates the principle of CSI-based behavior recognition. Secondly, the article presents the general framework of behavior recognition in detail, such as base signal selection, signal preprocessing, the behavior recognition techniques including pattern-based, model-based, and deep learning-based approaches. Thirdly, based on the above recognition techniques, the article categorizes the existing studies and applications into three groups and elaborates on each typical application from the test device, experimental scenarios, number of users, behaviors conducted, classifier and system performance. Fourthly, this paper analyzes some specific applications and presents extensive discussions on the selection of recognition techniques and performance evaluation. These discussions provide some helpful insights into developing an identification system. Finally, this article concludes by presenting open issues and research future.

REFERENCES

[1] J. Huang, N. Duan, P. Ji, C. Ma, F. Hu, Y. Ding, Y. Yu, Q. Zhou, and W. Sun, "A crowdsource-based sensing system for monitoring fine-grained air quality in urban environments," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 3240–3247, Apr. 2019.

[2] P. M. Santos, J. G. P. Rodrigues, S. B. Cruz, T. Lourenço, and P. M. d'Orey, Y. Luis, C. Rocha, S. Sousa, S. Crisóstomo, C. Queirós, S. Sargento, A. Aguiar, and J. Barros, "PortoLivingLab: An IoT-based sensing platform for smart cities," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 523–532, Apr. 2018.

[3] D. Wu, D. Zhang, C. Xu, H. Wang, and X. Li, "Device-free WiFi human sensing: From pattern-based to model-based approaches," *IEEE Commun. Mag.*, vol. 55, no. 10, pp. 91–97, Oct. 2017.

[4] J. Liu, Y. Wang, Y. Chen, J. Yang, X. Chen, and J. Cheng, "Tracking vital signs during sleep leveraging off-the-shelf WiFi," in *Proc. 16th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, Hangzhou, China, 2015, pp. 267–276.

[5] Z. Zhou, L. Shangguan, X. Zheng, L. Yang, and Y. Liu, "Design and implementation of an RFID-based customer shopping behavior mining system," *IEEE/ACM Trans. Netw.*, vol. 25, no. 4, pp. 2405–2418, Apr. 2017.

[6] Z. Yu, E. Xu, H. Du, B. Guo, and L. Yao, "Inferring user profile attributes from multidimensional mobile phone sensory data," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 5152–5162, Jun. 2019.

[7] X. Zheng, J. Wang, L. Shangguan, Z. Zhou, and Y. Liu, "Design and implementation of a CSI-based ubiquitous smoking detection system," *IEEE/ACM Trans. Netw.*, vol. 25, no. 6, pp. 3781–3793, Dec. 2017.

[8] H. Zou, Y. Zhou, J. Yang, H. Jiang, L. Xie, and C. J. Spanos, "WiFi-enabled device-free gesture recognition for smart home automation," in *Proc. IEEE 14th Int. Conf. Control Autom. (ICCA)*, Anchorage, AK, USA, Jun. 2018, pp. 476–481.

[9] B. Lashkari, J. Rezazadeh, R. Farahbakhsh, and K. Sandrasegaran, "Crowdsourcing and sensing for indoor localization in IoT: A review," *IEEE Sensors J.*, vol. 19, no. 7, pp. 2408–2434, Apr. 2019.

[10] K. Qian, C. Wu, Z. Yang, Y. Liu, and K. Jamieson, "Widar: Decimeter-level passive tracking via velocity monitoring with commodity Wi-Fi," in *Proc. 18th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, Chennai, India, 2017, pp. 1–10.

[11] 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 *Proc. 16th Annu. Int. Conf. Mobile Syst., Appl., Services*, Munich, Germany, 2018, pp. 350–361.

[12] J. Yang, H. Zou, H. Jiang, and L. Xie, "Device-free occupant activity sensing using WiFi-enabled IoT devices for smart homes," *IEEE Internet Things J.*, vol. 5, no. 5, pp. 3991–4002, Oct. 2018.

[13] C. S. Evangeline and A. Lenin, "Human health monitoring using wearable sensor," *Sensor Rev.*, vol. 39, no. 3, pp. 364–376, May 2019.

[14] Z. Yang, Z. Zhou, and Y. Liu, "From RSSI to CSI: Indoor localization via channel response," *ACM Comput. Surv.*, vol. 46, no. 2, pp. 1–32, 2013.

[15] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Predictable 802.11 packet delivery from wireless channel measurements," *ACM SIGCOM Comput. Commun. Rev.*, vol. 40, no. 4, pp. 159–170, Oct. 2010.

[16] J. Zhao, L. Liu, Z. Wei, C. Zhang, W. Wang, and Y. Fan, "R-DEHM: CSI-based robust duration estimation of human motion with WiFi," *Sensors*, vol. 19, no. 6, p. 1421, Mar. 2019.

[17] J.-S. Choi, W.-H. Lee, J.-H. Lee, J.-H. Lee, and S.-C. Kim, "Deep learning based NLOS identification with commodity WLAN devices," *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3295–3303, Apr. 2018.

[18] Y. Xie, Z. Li, and M. Li, "Precise power delay profiling with commodity Wi-Fi," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, Paris, France, 2015, pp. 53–64.

[19] S. Duan, T. Yu, and J. He, "WiDriver: Driver activity recognition system based on WiFi CSI," *Int. J. Wireless Inf. Netw.*, vol. 25, no. 2, pp. 146–156, Jan. 2018.

[20] S. Sen, J. Lee, K.-H. Kim, and P. Congdon, "Avoiding multipath to revive inbuilding WiFi localization," in *Proc. 11th Annu. Int. Conf. Mobile Syst., Appl., Services*, Taipei, Taiwan, 2013, pp. 249–262.

[21] H. Zou, Y. Zhou, J. Yang, W. Gu, L. Xie, and C. Spanos, "Poster: WiFi-based device-free human activity recognition via automatic representation learning," in *Proc. 23rd Annu. Int. Conf. Mobile Comput. Netw.*, Snowbird, UT, USA, 2017, pp. 606–608.

[22] U. M. Khan, Z. Kabir, S. A. Hassan, and S. H. Ahmed, "A deep learning framework using passive WiFi sensing for respiration monitoring," in *Proc. IEEE Global Commun. Conf.*, Singapore, Dec. 2017, pp. 1–6.

[23] Y. Xie, J. Xiong, M. Li, and K. Jamieson, "mD-Track: Leveraging multi-dimensionality in passive indoor Wi-Fi tracking," Dec. 2018, *arXiv:1812.03103*. [Online]. Available: https://arxiv.org/abs/1812.03103

[24] Y. Zeng, D. Wu, R. Gao, T. Gu, and D. Zhang, "FullBreathe: Full human respiration detection exploiting complementarity of CSI phase and amplitude of WiFi signals," *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.*, vol. 2, no. 3, pp. 1–148, Sep. 2018.

[25] M. Youssef, M. Mah, and A. Agrawala, "Challenges: Device-free passive localization for wireless environments," in *Proc. 13th Annu. ACM Int. Conf. Mobile Comput. Netw.*, Montreal, QC, Canada, 2007, pp. 222–229.

[26] J. Wilson and N. Patwari, "See-through walls: Motion tracking using variance-based radio tomography networks," *IEEE Trans. Mobile Comput.*, vol. 10, no. 5, pp. 612–621, May 2011.

[27] P. Xiang, P. Ji, and D. Zhang, "Enhance RSS-based indoor localization accuracy by leveraging environmental physical features," *Wireless Commun. Mobile Comput.*, vol. 2018, Jul. 2018, Art. no. 8956757. doi: 10.1155/2018/8956757.

[28] J. Wilson and N. Patwari, "Radio tomographic imaging with wireless networks," *IEEE Trans. Mobile Comput.*, vol. 9, no. 5, pp. 621–632, May 2010.

[29] J. Wang, X. Zhang, Q. Gao, H. Yue, and H. Wang, "Device-free wireless localization and activity recognition: A deep learning approach," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 6258–6267, Jul. 2017.

[30] M. A. A. Haseeb and R. Parasuraman, "Wisture: RNN-based learning of wireless signals for gesture recognition in unmodified smartphones," Jul. 2017, *arXiv:1707.08569*. [Online]. Available: https://arxiv.org/abs/1707.08569

[31] S. Lv, Y. Lu, M. Dong, X. Wang, Y. Dou, and W. Zhuang, "Qualitative action recognition by wireless radio signals in human–machine systems," *IEEE Trans. Human-Mach. Syst.*, vol. 47, no. 6, pp. 789–800, Dec. 2017.

[32] S. Depatla and Y. Mostofi, "Crowd counting through walls using WiFi," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. (PerCom)*, Athens, Greece, Mar. 2018, pp. 1–10.

[33] C. Xiang, Z. Zhang, S. Zhang, S. Xu, S. Cao, and V. Lau, "Robust sub-meter level indoor localization—A logistic regression approach," Feb. 2019, *arXiv:1902.06226*. [Online]. Available: https://arxiv.org/abs/1902.06226

[34] Y. Zheng, M. Sheng, J. Liu, and J. Li, "OpArray: Exploiting array orientation for accurate indoor localization," *IEEE Trans. Commun.*, vol. 67, no. 1, pp. 847–858, Jan. 2019.

[35] D. Zhu, B. Zhao, S. Wang, and D. Wu, "Direct-path based fingerprint extraction algorithm for indoor localization," in *Proc. 15th EAI Int. Conf. Mobile Ubiquitous Syst., Comput., Netw. Services (MobiQuitous)*, New York, NY, USA, 2018, pp. 11–18.

[36] L. Zheng, B. Hu, and H. Chen, "A high accuracy time-reversal based WiFi indoor localization approach with a single antenna," *Sensors*, vol. 18, no. 10, p. 3437, Oct. 2018.

[37] Y. Xiao, S. Zhang, J. Wang, and C. Zhu, "A novel indoor localization algorithm for efficient mobility management in wireless networks," *Wireless Commun. Mobile Comput.*, vol. 2018, Jun. 2018, Art. no. 9517942. doi: 10.1155/2018/9517942.

[38] Z. Wu, Q. Xu, J. Li, C. Fu, Q. Xuan, and Y. Xiang, "Passive indoor localization based on CSI and naive Bayes classification," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 48, no. 9, pp. 1566–1577, Sep. 2018.

[39] Y. Wang, C. Xiu, X. Zhang, and D. Yang, "WiFi indoor localization with CSI fingerprinting-based random forest," *Sensors*, vol. 18, no. 9, Aug. 2018, Art. no. 2869. doi: 10.3390/s18092869.

[40] F. Xiao, Z. Guo, H. Zhu, X. Xie, and R. Wang, "AmpN: Real-time LOS/NLOS identification with WiFi," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Paris, France, May 2017, pp. 1–7.

[41] X. Wang, L. Gao, and S. Mao, "BiLoc: Bi-modal deep learning for indoor localization with commodity 5GHz WiFi," *IEEE Access*, vol. 5, pp. 4209–4220, 2017.

[42] K. Ohara, T. Maekawa, and Y. Matsushita, "Detecting state changes of indoor everyday objects using Wi-Fi channel state information," in *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 1, no. 3, Sep. 2017, pp. 1–28.

[43] S. Zhong, Y. Huang, R. Ruby, L. Wang, Y. X. Qiu, and K. Wu, "Wi-fire: Device-free fire detection using WiFi networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Paris, France, May 2017, pp. 1–6.

[44] M. Won, S. Zhang, and S. H. Son, "WiTraffic: Low-cost and non-intrusive traffic monitoring system using WiFi," in *Proc. 26th Int. Conf. Comput. Commun. Netw. (ICCCN)*, Vancouver, BC, Canada, Jul. 2017, pp. 1–9.

[45] W. Yang, X. Wang, A. Song, and S. Mao, "Wi-wheat: Contact-free wheat moisture detection with commodity WiFi," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kansas City, MO, USA, May 2018, pp. 1–6.

[46] Y. Zou, Y. Wang, Y. Shufeng, K. Wu, and L. M. Ni, "TagFree: Passive object differentiation via physical layer radiometric signatures," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. (PerCom)*, Kona, HI, USA, Mar. 2017, pp. 237–246.

[47] C. Wang, J. Liu, Y. Chen, H. Liu, and Y. Wang, "Towards in-baggage suspicious object detection using commodity WiFi," in *Proc. IEEE Conf. Commun. Netw. Secur. (CNS)*, Beijing, China, May 2018, pp. 1–9.

[48] Q. Zhou, C. Wu, J. Xing, J. Li, Z. Yang, and Q. Yang, "Wi-dog: Monitoring school violence with commodity WiFi devices," in *Proc. 12th Int. Conf. Wireless Algorithms, Syst., Appl. (WASA)*, Guilin, China, Jun. 2017, pp. 47–59.
[49] S. Arshad, C. Feng, I. Elujide, S. Zhou, and Y. Liu, "SafeDrive-Fi: A multimodal and device free dangerous driving recognition system using WiFi," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kansas City, MO, USA, May 2018, pp. 1–6.
[50] H. Saeed, T. Saeed, M. Tahir, and M. Uppal, "Risky driving behavior detection using in-vehicle WiFi signals," in *Proc. IEEE 88th Veh. Technol. Conf. (VTC-Fall)*, Chicago, IL, USA, Aug. 2018, pp. 1–5.
[51] H. Peng and W. Jia, "WiFind: Driver fatigue detection with fine-grained Wi-Fi signal features," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Singapore, Dec. 2017, pp. 1–6.
[52] M. Raja, V. Ghaderi, and S. Sigg, "Detecting driver's distracted behaviour from Wi-Fi," in *Proc. IEEE 87th Veh. Technol. Conf. (VTC Spring)*, Porto, Portugal, Jun. 2018, pp. 1–5.
[53] M. Raja, V. Ghaderi, and S. Sigg, "WiBot! In-vehicle behaviour and gesture recognition using wireless network edge," in *Proc. IEEE 38th Int. Conf. Distrib. Comput. Syst. (ICDCS)*, Vienna, Austria, Jul. 2018, pp. 376–387.
[54] X. Xie, K. G. Shin, H. Yousefi, and S. He, "Wireless CSI-based head tracking in the driver seat," in *Proc. 14th Int. Conf. Emerg. Netw. Exp. Technol.*, Heraklion, Greece, Dec. 2018, pp. 112–125.
[55] 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," in *Proc. 20th Annu. Int. Conf. Mobile Comput. Netw.*, Maui, HI, USA, 2014, pp. 617–628.
[56] 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.*, Seattle, WA, USA, 2015, pp. 154–165.
[57] T. Z. Chowdhury, C. Leung, and C. Y. Miao, "WiHACS: Leveraging WiFi for human activity classification using OFDM subcarriers' correlation," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Montreal, QC, Canada, Nov. 2017, pp. 338–342.
[58] C. Feng, S. Arshad, and Y. Liu, "MAIS: Multiple activity identification system using channel state information of WiFi signals," in *Proc. 12th Int. Conf. Wireless Algorithms, Syst., Appl. (WASA)*, Guilin, China, Jun. 2017, pp. 419–432.
[59] R. H. Venkatnarayan, G. Page, and M. Shahzad, "Multi-user gesture recognition using WiFi," in *Proc. 16th Annu. Int. Conf. Mobile Syst., Appl., Services*, Munich, Germany, 2018, pp. 401–413.
[60] J. Wang, L. Zhang, Q. Gao, M. Pan, and H. Wang, "Device-free wireless sensing in complex scenarios using spatial structural information," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2432–2442, Apr. 2018.
[61] X. Wu, Z. Chu, P. Yang, C. Xiang, X. Zheng, and W. Huang, "TW-see: Human activity recognition through the wall with commodity Wi-Fi devices," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 306–319, Jan. 2019.
[62] S. Tan, L. Zhang, Z. Wang, and J. Yang, "MultiTrack: Multi-user tracking and activity recognition using commodity WiFi," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Glasgow, U.K., 2019, pp. 1–12.
[63] C. Chen, Y. Shu, K.-I. Shu, and H. Zhang, "WiTT: Modeling and the evaluation of table tennis actions based on WIFI signals," in *Proc. 24th Int. Conf. Pattern Recognit. (ICPR)*, Beijing, China, Aug. 2018, pp. 3100–3107.
[64] S. Li, X. Li, Q. Lv, G. Tian, and D. Zhang, "WiFit: Ubiquitous bodyweight exercise monitoring with commodity Wi-Fi devices," in *Proc. IEEE SmartWorld, Ubiquitous Intell. Comput., Adv. Trusted Comput., Scalable Comput. Commun., Cloud Big Data Comput., Internet People Smart City Innov. (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, Guangzhou, China, Oct. 2018, pp. 530–537.
[65] Z. Zhang, S. Ishida, S. Tagashira, and A. Fukuda, "Danger-pose detection system using commodity Wi-Fi for bathroom monitoring," *Sensors*, vol. 19, no. 4, Feb. 2019, Art. no. 884. doi: 10.3390/s19040884.
[66] D. Zhu, N. Pang, G. Li, and S. Liu, "NotiFi: A ubiquitous WiFi-based abnormal activity detection system," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Anchorage, AK, USA, 2017, pp. 1766–1773.
[67] 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. 2017.
[68] D. Zhang, H. Wang, Y. Wang, and J. Ma, "Anti-fall: A non-intrusive and real-time fall detector leveraging CSI from commodity WiFi devices," in *Proc. 13th Int. Conf. Smart Homes Health Telematics, Inclusive Smart Cities e-Health (ICOST)*, Geneva, Switzerland, Jun. 2015, pp. 181–193.
[69] H. Wang, D. Zhang, Y. Wang, J. Ma, Y. Wang, and S. Li, "RT-Fall: A real-time and contactless fall detection system with commodity WiFi devices," *IEEE Trans. Mobile Comput.*, vol. 16, no. 2, pp. 511–526, Feb. 2017.
[70] S. Palipana, D. Rojas, P. Agrawal, and D. Pesch, "FallDeFi: Ubiquitous fall detection using commodity Wi-Fi devices," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 1, no. 4, pp. 1–25, Dec. 2018.
[71] Z. Dong, F. Li, J. Ying, and K. Pahlavan, "Indoor motion detection using Wi-Fi channel state information in flat floor environments versus in staircase environments," *Sensors*, vol. 18, no. 7, Jul. 2018, Art. no. 2177. doi: 10.3390/s18072177.
[72] R. Nandakumar, B. Kellogg, and S. Gollakota, "Wi-Fi gesture recognition on existing devices," Nov. 2014, *arXiv:1411.5394*. [Online]. Available: https://arxiv.org/abs/1411.5394
[73] W. He, K. Wu, Y. Zou, and Z. Ming, "WiG: WiFi-based gesture recognition system," in *Proc. 24th Int. Conf. Comput. Commun. Netw. (ICCCN)*, Las Vegas, NV, USA, Aug. 2015, pp. 1–7.
[74] M. Al-qaness and F. Li, "WiGeR: WiFi-based gesture recognition system," *ISPRS Int. J. Geo-Inf.*, vol. 5, no. 6, Jun. 2016, Art. no. 92. doi: 10.3390/ijgi5060092.
[75] O. Zhang and K. Srinivasan, "Mudra: User-friendly fine-grained gesture recognition using WiFi signals," in *Proc. 12th Int. Conf. Emerg. Netw. Exp. Technol.*, Irvine, CA, USA, 2016, pp. 83–96.
[76] H. Li, W. Yang, J. Wang, Y. Xu, and L. Huang, "WiFinger: Talk to your smart devices with finger-grained gesture," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, Heidelberg, Germany, 2016, pp. 250–261.
[77] S. Tan and J. Yang, "WiFinger: Leveraging commodity WiFi for fine-grained finger gesture recognition," in *Proc. 17th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, Paderborn, Germany, 2016, pp. 201–210.
[78] D. Man, W. Yang, X. Wang, J. Lv, X. Du, and M. Yu, "PWiG: A phase-based wireless gesture recognition system," in *Proc. Int. Conf. Comput., Netw. Commun. (ICNC)*, Maui, HI, USA, 2018, pp. 837–842.
[79] J. Chen, F. Li, H. Chen, S. Yang, and Y. Wang, "Dynamic gesture recognition using wireless signals with less disturbance," *Pers. Ubiquitous Comput.*, vol. 23, no. 1, pp. 17–27, Feb. 2019.
[80] Z. Tian, J. Wang, X. Yang, and M. Zhou, "WiCatch: A Wi-Fi based hand gesture recognition system," *IEEE Access*, vol. 6, pp. 16911–16923, Apr. 2018.
[81] S. Ren, H. Wang, L. Gong, C. Xiang, X. Wu, and Y. Du, "Intelligent contactless gesture recognition using WLAN physical layer information," *IEEE Access*, vol. 7, pp. 92758–92767, Jul. 2019.
[82] J. Shang and J. Wu, "A robust sign language recognition system with multiple Wi-Fi devices," in *Proc. Workshop Mobility Evolving Internet Archit.*, Los Angeles, CA, USA, 2017, pp. 19–24.
[83] J. Shang and J. Wu, "A robust sign language recognition system with sparsely labeled instances using Wi-Fi signals," in *Proc. IEEE 14th Int. Conf. Mobile Ad Hoc Sensor Syst. (MASS)*, Orlando, FL, USA, Oct. 2017, pp. 99–107.
[84] X. Liu, J. Cao, S. Tang, and J. Wen, "Wi-sleep: Contactless sleep monitoring via WiFi signals," in *Proc. IEEE Real-Time Syst. Symp. (RTSS)*, Rome, Italy, Dec. 2014, pp. 346–355.
[85] X. Wang, C. Yang, and S. Mao, "PhaseBeat: Exploiting CSI phase data for vital sign monitoring with commodity WiFi devices," in *Proc. IEEE 37th Int. Conf. Distrib. Comput. Syst. (ICDCS)*, Atlanta, GA, USA, Jun. 2017, pp. 1230–1239.
[86] X. Wang, C. Yang, and S. Mao, "TensorBeat: Tensor decomposition for monitoring multiperson breathing beats with commodity WiFi," *ACM Trans. Intell. Syst. Technol.*, vol. 9, no. 1, pp. 1–27, Sep. 2017.
[87] C. Chen, Y. Han, Y. Chen, H. Q. Lai, F. Zhang, B. Wang, and K. J. R. Liu, "TR-BREATH: Time-reversal breathing rate estimation and detection," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 3, pp. 489–501, Mar. 2018.
[88] J. Liu, Y. Chen, Y. Wang, X. Chen, J. Cheng, and J. Yang, "Monitoring vital signs and postures during sleep using WiFi signals," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 2071–2084, Jun. 2018.
[89] G. Wang, Y. Zou, Z. Zhou, K. Wu, and L. M. Ni, "We can hear you with Wi-Fi!" *IEEE Trans. Mobile Comput.*, vol. 15, no. 11, pp. 2907–2920, Nov. 2016.
[90] C. Du, X. Yuan, W. Lou, and Y. T. Hou, "Context-free fine-grained motion sensing using WiFi," in *Proc. 15th Annu. IEEE Int. Conf. Sens., Commun., Netw. (SECON)*, Hong Kong, Jun. 2018, pp. 1–9.

[91] K. Ali, A. X. Liu, W. Wang, and M. Shahzad, "Keystroke recognition using WiFi signals," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, Paris, France, 2015, pp. 90–102.

[92] M. Li, Y. Meng, J. Liu, H. Zhu, X. Liang, Y. Liu, and N. Ruan, "When CSI meets public WiFi: Inferring your mobile phone password via WiFi signals," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Vienna, Austria, 2016, pp. 1068–1079.

[93] X. Cao, B. Chen, and Y. Zhao, "Wi-Wri: Fine-grained writing recognition using Wi-Fi signals," in *Proc. IEEE Trustcom/BigDataSE/ISPA*, Tianjin, China, Aug. 2016, pp. 1366–1373.

[94] L. Zhang, J. Wang, Q. Gao, X. Li, M. Pan, and Y. Fang, "LetFi: Letter recognition in the air using CSI," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Abu Dhabi, United Arab Emirates, Dec. 2018, pp. 1–6.

[95] Z. Fu, J. Xu, Z. Zhu, A. X. Liu, and X. Sun, "Writing in the air with WiFi signals for virtual reality devices," *IEEE Trans. Mobile Comput.*, vol. 18, no. 2, pp. 473–484, Feb. 2019.

[96] J. Yang, H. Zou, H. Jiang, and L. Xie, "CareFi: Sedentary behavior monitoring system via commodity WiFi infrastructures," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 7620–7629, Aug. 2018.

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

[98] W. Xi, J. Zhao, X. Y. Li, K. Zhao, S. Tang, X. Liu, and Z. Jiang, "Electronic frog eye: Counting crowd using WiFi," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Toronto, ON, Canada, Apr. 2014, pp. 361–369.

[99] H. Zou, Y. Zhou, J. Yang, W. Gu, L. Xie, and C. Spanos, "FreeCount: Device-free crowd counting with commodity WiFi," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Singapore, Dec. 2017, pp. 1–6.

[100] X. Guo, B. Liu, C. Shi, H. Liu, Y. Chen, and M. C. Chuah, "WiFi-enabled smart human dynamics monitoring," in *Proc. 15th ACM Conf. Embedded Netw. Sensor Syst.*, Delft, The Netherlands, 2017, pp. 1–13.

[101] H. Zou, Y. Zhou, J. Yang, and C. J. Spanos, "Device-free occupancy detection and crowd counting in smart buildings with WiFi-enabled IoT," *Energy Buildings*, vol. 174, pp. 309–322, Sep. 2018.

[102] Y. Yang, J. Cao, X. Liu, and X. Liu, "Wi-count: Passing people counting with COTS WiFi devices," in *Proc. 27th Int. Conf. Comput. Commun. Netw. (ICCCN)*, Hangzhou, China, Jul. 2018, pp. 1–9.

[103] F. Xiao, Z. Guo, Y. Ni, X. Xie, S. Maharjan, and Y. Zhang, "Artificial intelligence empowered mobile sensing for human flow detection," *IEEE Netw.*, vol. 33, no. 1, pp. 78–83, Jan./Feb. 2019.

[104] L. Zhang, M. Liu, L. Lu, and L. Gong, "Wi-run: Multi-runner step estimation using commodity Wi-Fi," in *Proc. 15th Annu. IEEE Int. Conf. Sens., Commun., Netw. (SECON)*, Hong Kong, Jun. 2018, pp. 1–9.

[105] Y. Xu, W. Yang, J. Wang, X. Zhou, H. Li, and L. Huang, "WiStep: Device-free step counting with WiFi signals," *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.*, vol. 1, no. 4, pp. 1–23, Jan. 2018.

[106] C. Wu, Z. Yang, Z. Zhou, X. Liu, Y. Liu, and J. Cao, "Non-invasive detection of moving and stationary human with WiFi," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 11, pp. 2329–2342, Nov. 2015.

[107] E. Soltanaghaei, A. Kalyanaraman, and K. Whitehouse, "Peripheral WiFi vision: Exploiting multipath reflections for more sensitive human sensing," in *Proc. 4th Int. Workshop Phys. Anal.*, Niagara Falls, NY, USA, 2017, pp. 13–18.

[108] H. Zhu, F. Xiao, L. Sun, R. Wang, and P. Yang, "R-TTWD: Robust device-free through-the-wall detection of moving human with WiFi," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 5, pp. 1090–1103, May 2017.

[109] J. Lv, D. Man, W. Yang, X. Du, and M. Yu, "Robust WLAN-based indoor intrusion detection using PHY layer information," *IEEE Access*, vol. 6, pp. 30117–30127, Jun. 2018.

[110] K. Qian, C. Wu, Z. Yang, Y. Liu, F. He, and T. Xing, "Enabling contactless detection of moving humans with dynamic speeds using CSI," *ACM Trans. Embed. Comput. Syst.*, vol. 17, no. 2, pp. 1–18, Jan. 2018.

[111] S. D. Domenico, M. D. Sanctis, E. Cianca, and M. Ruggieri, "WiFi-based through-the-wall presence detection of stationary and moving humans analyzing the Doppler spectrum," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 33, nos. 5–6, pp. 14–19, May 2018.

[112] J. Zhang, B. Wei, W. Hu, and S. S. Kanhere, "WiFi-ID: Human identification using WiFi signal," in *Proc. Int. Conf. Distrib. Comput. Sensor Syst. (DCOSS)*, Washington, DC, USA, May 2016, pp. 75–82.

[113] Y. Zeng, P. H. Pathak, and P. Mohapatra, "WiWho: WiFi-based person identification in smart spaces," in *Proc. 15th ACM/IEEE Int. Conf. Inf. Process. Sensor Netw. (IPSN)*, Vienna, Austria, Apr. 2016, pp. 1–12.

[114] W. Wang, A. X. Liu, and M. Shahzad, "Gait recognition using WiFi signals," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, Heidelberg, Germany, 2016, pp. 363–373.

[115] T. Xin, B. Guo, Z. Wang, M. Li, Z. Yu, and X. Zhou, "FreeSense: Indoor human identification with Wi-Fi signals," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Washington, DC, USA, Dec. 2016, pp. 1–7.

[116] J. Lv, W. Yang, and D. Man, "Device-free passive identity identification via WiFi signals," *Sensors*, vol. 17, no. 11, Nov. 2017, Art. no. 2520. doi: 10.3390/s17112520.

[117] Y. Chen, W. Dong, Y. Gao, X. Liu, and T. Gu, "Rapid: A multimodal and device-free approach using noise estimation for robust person identification," *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.*, vol. 1, no. 3, pp. 1–27, Sep. 2017.

[118] H. Zou, Y. Zhou, J. Yang, W. Gu, L. Xie, and C. Spanos, "WiFi-based human identification via convex tensor shapelet learning," in *Proc. AAAI Conf. Artif. Intell.*, New Orleans, LA, USA, 2018, pp. 1711–1718.

[119] S. W. Shah and S. S. Kanhere, "Wi-sign: Device-free second factor user authentication," in *Proc. 15th EAI Int. Conf. Mobile Ubiquitous Syst., Comput., Netw. Services*, New York, NY, USA, 2018, pp. 135–144.

[120] H. Liu, Y. Wang, J. Liu, J. Yang, Y. Chen, and H. V. Poor, "Authenticating users through fine-grained channel information," *IEEE Trans. Mobile Comput.*, vol. 17, no. 2, pp. 251–264, Feb. 2018.

[121] F. Wang, Z. Li, and J. Han, "Continuous user authentication by contactless wireless sensing," Dec. 2018, *arXiv:1812.01503*. [Online]. Available: https://arxiv.org/abs/1812.01503

[122] Z. Zhao, Z. Zhao, G. Min, C. Shu, Z. Wang, and H. Duan, "Non-intrusive biometric identification for personalized computing using wireless big data," in *Proc. IEEE SmartWorld, Ubiquitous Intell. Comput., Adv. Trusted Comput., Scalable Comput. Commun., Cloud Big Data Comput., Internet People Smart City Innov. (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, Guangzhou, China, Oct. 2018, pp. 901–908.

[123] J. Liu, Y. Dong, Y. Chen, Y. Wang, and T. Zhao, "Poster: Leveraging breathing for continuous user authentication," in *Proc. 24th Annu. Int. Conf. Mobile Computing Netw.*, New Delhi, India, 2018, pp. 786–788.

[124] 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.*, Paris, France, 2015, pp. 65–76.

[125] 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.

[126] T. Xin, B. Guo, Z. Wang, P. Wang, J. C. K. Lam, V. Li, and Z. Yu, "FreeSense: A robust approach for indoor human detection using Wi-Fi signals," *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.*, vol. 2, no. 3, pp. 1–23, Sep. 2018.

[127] J. Lv, D. Man, W. Yang, L. Gong, X. Du, and M. Yu, "Robust device-free intrusion detection using physical layer information of WiFi signals," *Appl. Sci.*, vol. 9, no. 1, Jan. 2019, Art. no. 175. doi: 10.3390/app9010175.

[128] F. Wang, J. Han, Z. Dai, H. Ding, and D. Huang, "WiPIN: Operation-free person identification using WiFi signals," Oct. 2018, *arXiv:1810.04106*. [Online]. Available: https://arxiv.org/abs/1810.04106

[129] D. Wu, D. Zhang, C. Xu, Y. Wang, and H. Wang, "WiDir: Walking direction estimation using wireless signals," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, Heidelberg, Germany, 2016, pp. 351–362.

[130] L. Sun, S. Sen, D. Koutsonikolas, and K.-H. Kim, "WiDraw: Enabling hands-free drawing in the air on commodity WiFi devices," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, Paris, France, 2015, pp. 77–89.

[131] H. Wang, D. Zhang, J. Ma, Y. Wang, Y. Wang, D. Wu, T. Gu, and B. Xie, "Human respiration detection with commodity WiFi devices: Do user location and body orientation matter?" in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, Heidelberg, Germany, 2016, pp. 25–36.

[132] P. Wang, B. Guo, T. Xin, Z. Wang, and Z. Yu, "TinySense: Multi-user respiration detection using Wi-Fi CSI signals," in *Proc. IEEE 19th Int. Conf. e-Health Netw., Appl. Services (Healthcom)*, Dalian, China, Oct. 2017, pp. 1–6.

[133] K. Niu, F. Zhang, Z. Chang, and D. Zhang, "A Fresnel diffraction model based human respiration detection system using COTS Wi-Fi devices," in *Proc. ACM Int. Joint Conf. Int. Symp. Pervasive Ubiquitous Comput. Wearable Comput.*, Singapore, 2018, pp. 416–419.

[134] Y. Yang, J. Cao, X. Liu, and K. Xing, "Multi-person sleeping respiration monitoring with COTS WiFi devices," in *Proc. IEEE 15th Int. Conf. Mobile Ad Hoc Sensor Syst. (MASS)*, Chengdu, China, Oct. 2018, pp. 37–45.

[135] F. Zhang, D. Zhang, J. Xiong, H. Wang, K. Niu, B. Jin, and Y. Wang, "From Fresnel diffraction model to fine-grained human respiration sensing with commodity Wi-Fi devices," *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.*, vol. 2, no. 1, pp. 1–23, Mar. 2018.
[136] D. Zhang, Y. Hu, Y. Chen, and B. Zeng, "BreathTrack: Tracking indoor human breath status via commodity WiFi," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 3899–3911, Apr. 2019.
[137] Y. Zeng, D. Wu, J. Xiong, E. Yi, R. Gao, and D. Zhang, "FarSense: Pushing the range limit of WiFi-based respiration sensing with CSI ratio of two antennas," Jul. 2019, *arXiv:1907.03994*. [Online]. Available: https://arxiv.org/abs/1907.03994
[138] F. X. Wang, W. Gong, and J. C. Liu, "On spatial diversity in WiFi-based human activity recognition: A deep learning-based approach," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2035–2047, Apr. 2019.
[139] H. Zou, Y. Zhou, J. Yang, and C. J. Spanos, "Towards occupant activity driven smart buildings via WiFi-enabled IoT devices and deep learning," *Energy Buildings*, vol. 177, pp. 12–22, Oct. 2018.
[140] H. Zou, Y. Zhou, J. Yang, H. Jiang, L. Xie, and C. J. Spanos, "DeepSense: Device-free human activity recognition via autoencoder long-term recurrent convolutional network," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kansas City, MO, USA, May 2018, pp. 1–6.
[141] D. A. Khan, S. Razak, B. Raj, and R. Singh, "Human behaviour recognition using Wifi channel state information," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Brighton, U.K., May 2019, pp. 7625–7629.
[142] F. Wang, S. Panev, Z. Dai, J. Han, and D. Huang, "Can WiFi estimate person pose?" Mar. 2019, *arXiv:1904.00277*. [Online]. Available: https://arxiv.org/abs/1904.00277
[143] T. Z. Chowdhury, "Using Wi-Fi channel state information (CSI) for human activity recognition and fall detection," Ph.D. dissertation, Dept. Elect. Comput. Eng., Univ. British Columbia, Vancouver, BC Canada, 2018.
[144] X. Yang, F. Xiong, Y. Shao, and Q. Niu, "WmFall: WiFi-based multistage fall detection with channel state information," *Int. J. Distrib. Sensor Netw.*, vol. 14, no. 10, pp. 1–10, Oct. 2018.
[145] Z. Wang, Z. Gu, J. Yin, Z. Chen, and Y. Xu, "Syncope detection in toilet environments using Wi-Fi channel state information," in *Proc. ACM Int. Joint Conf. Int. Symp. Pervasive Ubiquitous Comput. Wearable Comput.*, Singapore, 2018, pp. 287–290.
[146] Q. Zhou, J. Xing, J. Li, and Q. Yang, "A device-free number gesture recognition approach based on deep learning," in *Proc. 12th Int. Conf. Comput. Intell. Secur. (CIS)*, Wuxi, China, Dec. 2016, pp. 57–63.
[147] Q. Zhou, J. Xing, W. Chen, X. Zhang, and Q. Yang, "From signal to image: Enabling fine-grained gesture recognition with commercial Wi-Fi devices," *Sensors*, vol. 18, no. 9, p. 3142, 2018.
[148] Y. Zheng, Y. Zhang, K. Qian, G. Zhang, Y. Liu, C. Wu, and Z. Yang, "Zero-effort cross-domain gesture recognition with Wi-Fi," in *Proc. 17th Annu. Int. Conf. Mobile Syst., Appl., Services*, Seoul, Republic Korea, 2019, pp. 313–325.
[149] F. Wang, Y. Song, J. Zhang, J. Han, and D. Huang, "Temporal Unet: Sample level human action recognition using WiFi," Apr. 2019, *arXiv:1904.11953*. [Online]. Available: https://arxiv.org/abs/1904.11953
[150] Y. Ma, G. Zhou, S. Wang, H. Zhao, and W. Jung, "SignFi: Sign language recognition using WiFi," *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.*, vol. 2, no. 1, pp. 1–21, Mar. 2018.
[151] Y. Xu, M. Chen, W. Yang, S. Chen, and L. Huang, "Attention-based walking gait and direction recognition in Wi-Fi networks," Nov. 2018, *arXiv:1811.07162*. [Online]. Available: https://arxiv.org/abs/1811.07162
[152] K. Jakkala, A. Bhuya, Z. Sun, P. Wang, and Z. Cheng, "Deep CSI learning for gait biometric sensing and recognition," Feb. 2019, *arXiv:1902.02300*. [Online]. Available: https://arxiv.org/abs/1902.02300
[153] F. Wang, S. Zhou, S. Panev, J. Han, and D. Huang, "Person-in-WiFi: Fine-grained person perception using WiFi," Mar. 2019, *arXiv:1904.00276*. [Online]. Available: https://arxiv.org/abs/1904.00276
[154] F. Wang, J. Han, S. Zhang, X. He, and D. Huang, "CSI-Net: Unified human body characterization and action recognition," Oct. 2018, *arXiv:1810.03064*. [Online]. Available: https://arxiv.org/abs/1810.03064
[155] S.-H. Fang, C.-C. Li, W.-C. Lu, Z. Xu, and Y.-R. Chien, "Enhanced device-free human detection: Efficient learning from phase and amplitude of channel state information," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 3048–3051, Mar. 2019.
[156] Y.-K. Cheng and R. Y. Chang, "Device-free indoor people counting using Wi-Fi channel state information for Internet of Things," in *Proc. IEEE Globecom*, Singapore, Dec. 2017, pp. 1–6.
[157] S. Liu, Y. Zhao, and B. Chen, "WiCount: A deep learning approach for crowd counting using WiFi signals," in *Proc. IEEE Int. Symp. Parallel Distrib. Process. Appl. IEEE Int. Conf. Ubiquitous Comput. Commun. (ISPA/IUCC)*, Guangzhou, China, Dec. 2017, pp. 967–974.
[158] S. Doong, "Counting human flow with deep neural network," in *Proc. 51st Hawaii Int. Conf. Syst. Sci.*, Big Island, HI, USA, 2018, pp. 799–808.
[159] S. Liu, Y. Zhao, F. Xue, B. Chen, and X. Chen, "DeepCount: Crowd counting with WiFi via deep learning," Mar. 2019, *arXiv:1903.05316*. [Online]. Available: https://arxiv.org/abs/1903.05316
[160] C. Shi, J. Liu, H. Liu, and Y. Chen, "Smart user authentication through actuation of daily activities leveraging WiFi-enabled IoT," in *Proc. 18th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, Chennai, India, 2017, pp. 1–10.
[161] C. Lin, J. Hu, Y. Sun, F. Ma, L. Wang, and G. Wu, "WiAU: An accurate device-free authentication system with ResNet," in *Proc. 15th Annu. IEEE Int. Conf. Sens., Commun., Netw. (SECON)*, Hong Kong, Jun. 2018, pp. 1–9.
[162] A. Pokkunuru, K. Jakkala, A. Bhuyan, P. Wang, and Z. Sun, "NeuralWave: Gait-based user identification through commodity WiFi and deep learning," in *Proc. 44th Annu. Conf. IEEE Ind. Electron. Soc. (IECON)*, Washington, DC, USA, Oct. 2018, pp. 758–765.
[163] H. Kong, L. Lu, J. Yu, Y. Chen, L. Kong, and M. Li, "FingerPass: Finger gesture-based continuous user authentication for smart homes using commodity WiFi," in *Proc. 12th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, Catania, Italy, 2019, pp. 201–210.
[164] J. Ma, H. Wang, D. Zhang, Y. Wang, and Y. Wang, "A survey on Wi-Fi based contactless activity recognition," in *Proc. Int. IEEE Conf. Ubiquitous Intell. Comput., Adv. Trusted Comput., Scalable Comput. Commun., Cloud Big Data Comput., Internet People, Smart World Congr. (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld)*, Toulouse, France, Jul. 2016, pp. 1086–1091.
[165] 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.
[166] Z. Wang, B. Guo, Z. Yu, and X. Zhou, "Wi-Fi CSI-based behavior recognition: From signals and actions to activities," *IEEE Commun. Mag.*, vol. 56, no. 5, pp. 109–115, May 2018.
[167] H. Jiang, C. Cai, X. Ma, Y. Yang, and J. Liu, "Smart home based on WiFi sensing: A survey," *IEEE Access*, vol. 6, pp. 13317–13325, Mar. 2018.
[168] J. Wang, Q. Gao, M. Pan, and Y. Fang, "Device-free wireless sensing: Challenges, opportunities, and applications," *IEEE Netw.*, vol. 32, no. 2, pp. 132–137, Mar./Apr. 2018.
[169] Y. Lu, S.-H. Lv, X.-D. Wang, and X.-M. Zhou, "A Survey on WiFi based human behavior analysis technology," *Chin. J. Comput.*, vol. 41, no. 27, pp. 1–22, Mar. 2018.
[170] A. M. Khalili, A.-H. Soliman, M. Asaduzzaman, and A. Griffiths, "Wi-Fi sensing: Applications and challenges," Jan. 2019, *arXiv:1901.00715*. [Online]. Available: https://arxiv.org/abs/1901.00715
[171] Y. Ma, G. Zhou, and S. Wang, "WiFi sensing with channel state information: A survey," *ACM Comput. Surv.*, vol. 52, no. 3, pp. 1–36, Jul. 2019.
[172] Z. Wang, K. Jiang, Y. Hou, Z. Huang, W. Dou, C. Zhang, and Y. Guo, "A survey on CSI-based human behavior recognition in through-the-wall scenario," *IEEE Access*, vol. 7, pp. 78772–78793, Jun. 2019.
[173] J. Liu, H. Liu, Y. Chen, Y. Wang, and C. Wang, "Wireless sensing for human activity: A survey," *IEEE Commun. Surveys Tuts.*, to be published. doi: 10.1109/COMST.2019.2934489.
[174] T. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2001, p. 107.
[175] P. Melgarejo, X. Zhang, P. Ramanathan, and D. Chu, "Leveraging directional antenna capabilities for fine-grained gesture recognition," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, Seattle, WA, USA, 2014, pp. 541–551.
[176] D. Zhu, N. Pang, G. Li, and S. Liu, "WiseFi: Activity localization and recognition on commodity off-the-shelf WiFi devices," in *Proc. IEEE 18th Int. Conf. High Perform. Comput. Commun., IEEE 14th Int. Conf. Smart City, IEEE 2nd Int. Conf. Data Sci. Syst. (HPCC/SmartCity/DSS)*, Sydney, NSW, Australia, Dec. 2016, pp. 562–569.
[177] X. Liu, J. Cao, S. Tang, J. Wen, and P. Guo, "Contactless respiration monitoring via off-the-shelf WiFi devices," *IEEE Trans. Mobile Comput.*, vol. 15, no. 10, pp. 2466–2479, Oct. 2016.
[178] C. Feng, S. Arshad, R. Yu, and Y. Liu, "Evaluation and improvement of activity detection systems with recurrent neural network," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kansas City, MO, USA, May 2018, pp. 1–6.

[179] F. Zhang, C. Chen, B. B. Wang, and K. J. R. Liu, "WiSpeed: A statistical electromagnetic approach for device-free indoor speed estimation," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 2163–2177, Jun. 2018.

[180] J. Shang and J. Wu, "Fine-grained vital signs estimation using commercial Wi-Fi devices," in *Proc. 8th Wireless Students Workshop*, New York, NY, USA, 2016, pp. 30–32.

[181] S. Sigg, S. Shi, and Y. Ji, "RF-based device-free recognition of simultaneously conducted activities," in *Proc. ACM Conf. Pervasive Ubiquitous Comput. Adjunct Publication*, Zurich, Switzerland, 2013, pp. 531–540.

[182] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-home gesture recognition using wireless signals," in *Proc. 19th Annu. Int. Conf. Mobile Comput. Netw.*, Miami, FL, USA, 2013, pp. 27–38.

[183] S. D. Domenico, M. D. Sanctis, E. Cianca, F. Giuliano, and G. Bianchi, "Exploring training options for RF sensing using CSI," *IEEE Commun. Mag.*, vol. 56, no. 5, pp. 116–123, May 2018.

[184] R. Nandakumar, S. Gollakota, and N. Watson, "Contactless sleep apnea detection on smartphones," in *Proc. 13th Annu. Int. Conf. Mobile Syst., Appl., Services*, Florence, Italy, 2015, pp. 45–57.

[185] B. Tan, Q. Chen, K. Chetty, K. Woodbridge, W. Li, and R. Piechocki, "Exploiting WiFi channel state information for residential healthcare informatics," *IEEE Commun. Mag.*, vol. 56, no. 5, pp. 130–137, May 2018.

[186] Y. Zou, W. Liu, K. Wu, and L. M. Ni, "Wi-Fi radar: Recognizing human behavior with commodity Wi-Fi," *IEEE Commun. Mag.*, vol. 55, no. 10, pp. 105–111, Oct. 2017.

[187] Q. Bu, G. Yang, J. Feng, and X. Ming, "Wi-Fi based gesture recognition using deep transfer learning," in *Proc. IEEE SmartWorld, Ubiquitous Intell. Comput., Adv. Trusted Comput., Scalable Comput. Commun., Cloud Big Data Comput., Internet People Smart City Innov. (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, Guangzhou, China, Oct. 2018, pp. 590–595.

[188] 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.*, New Delhi, India, 2018, pp. 289–304.

[189] Q. Zhao, "A review of principal component analysis," *Softw. Eng.*, vol. 19, no. 6, pp. 1–3, Jun. 2016.

[190] Q. Kun, W. Chenshu, Y. Zheng, L. Yunhao, and Z. Zimu, "PADS: Passive detection of moving targets with dynamic speed using PHY layer information," in *Proc. 20th IEEE Int. Conf. Parallel Distrib. Syst. (ICPADS)*, Hsinchu, Taiwan, Dec. 2014, pp. 1–8.

[191] C. Wen, "Related issues of feature extraction and representation in behavior recognition," Ph.D. dissertation, Dept. Comput. Eng., Jilin Univ., Changchun, China, 2017.

[192] H. Wang, D. Zhang, K. Niu, Q. Lv, Y. Liu, D. Wu, R. Gao, and B. Xie, "MFDL: A multicarrier Fresnel penetration model based device-free localization system leveraging commodity Wi-Fi cards," Jul. 2017, *arXiv:1707.07514*. [Online]. Available: https://arxiv.org/abs/1707.07514

[193] D. Zhang, H. Wang, and D. Wu, "Toward centimeter-scale human activity sensing with Wi-Fi signals," *Computer*, vol. 50, no. 1, pp. 48–57, Jan. 2017.

[194] L. Chen, X. Chen, L. Ni, Y. Peng, and D. Fang, "Human behavior recognition using Wi-Fi CSI: Challenges and opportunities," *IEEE Commun. Mag.*, vol. 55, no. 10, pp. 112–117, Oct. 2017.

[195] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[196] J. Huang, B. Liu, H. Jin, and Z. Liu, "WiAnti: An anti-interference activity recognition system based on WiFi CSI," in *Proc. IEEE Int. Conf. Internet Things (iThings) IEEE Green Comput. Commun. (GreenCom) IEEE Cyber, Phys. Social Comput. (CPSCom) IEEE Smart Data (SmartData)*, Halifax, NS, Canada, Jul. 2018, pp. 58–65.

[197] C. Chen, Y. Han, Y. Chen, and K. J. R. Liu, "Multi-person breathing rate estimation using time-reversal on WiFi platforms," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Washington, DC, USA, Dec. 2016, pp. 1059–1063.

[198] H. Li, K. Ota, M. Dong, and M. Guo, "Learning human activities through Wi-Fi channel state information with multiple access points," *IEEE Commun. Mag.*, vol. 56, no. 5, pp. 124–129, May 2018.

[199] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," Dept. Comput. Sci., Univ. Toronto, Toronto, ON, Canada, Tech. Rep., Apr. 2009. [Online]. Available: http://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf

[200] J. Deng, W. Dong, R. Socher, L. Li, L. Kai, and F.-F. Li, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Miami, FL, USA, Jun. 2009, pp. 248–255.

[201] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[202] J. Zhang, Z. Tang, M. Li, D. Fang, X. Chen, and Z. Wang, "Find me a safe zone: A countermeasure for channel state information based attacks," *Comput. Secur.*, vol. 80, pp. 273–290, Jan. 2019.

ZHENGJIE WANG was born in Liaoyang, Liaoning, China, in 1972. He received the B.S. degree in power engineering from the North China University of Water Resources and Electric Power, Henan, China, in 1995, the M.S. degree in computer software and theory from Northeast University, Liaoning, China, in 2003, and the Ph.D. degree in computer application technology from the China University of Mining and Technology at Beijing, Beijing, China, in 2013.

Since 2003, he has been a Lecturer with the College of Electronic and Information Engineering, Shandong University of Science and Technology. He is the author of two books and more than ten articles. His research interests include human behavior recognition, people activity inference, person counting, identity authentication, and people tracking using WiFi devices and smartphones.

KANGKANG JIANG was born in Jining, Shandong, China, in 1993. She received the B.S. degree from Jining University, in 2017. She is currently pursuing the M.S. degree in communication and information system with the Shandong University of Science and Technology. Her research interests include image processing, machine learning, and deep learning.

YUSHAN HOU was born in Jinan, Shandong, China, in 1995. She received the B.S. degree from the Shandong University of Science and Technology, in 2017, where she is currently pursuing the M.S. degree in communication and information system. Her research interests include deep learning, machine learning, and signal processing.

WENWEN DOU was born in Taian, Shandong, China, in 1996. She received the B.S. degree from Weifang University, in 2018. She is currently pursuing the M.S. degree in communication and information system with the Shandong University of Science and Technology. Her research interests include deep learning, machine learning, and signal processing.

CHENGMING ZHANG was born in Zaozhuang, Shandong, China, in 1995. He received the B.S. degree from the Qilu University of Technology, in 2018. He is currently pursuing the M.S. degree in electronic and communication engineering with the Shandong University of Science and Technology. His research interests include machine learning, deep learning, and signal processing.

ZEHUA HUANG was born in Binzhou, Shandong, China, in 1996. He received the B.S. degree from the Shandong University of Science and Technology, in 2018, where he is currently pursuing the M.S. degree in communication and information system. His research interests include machine learning, deep learning, and signal processing.

YINJING GUO was born in Jining, Shandong, China, in 1966. He received the B.S. degree in radar engineering from the Ordnance Engineering College, in 1989, and the M.S. degree in communication and electronic system and the Ph.D. degree in weapon system and application engineering from the Beijing Institute of Technology, in 1992 and 2004, respectively.

Since 1996, he has been a Professor with the College of Electronic and Information Engineering, Shandong University of Science and Technology. He has published more than 90 articles, among which more than 40 articles have been retrieved by EI and SCI. His research interests include wireless communications, electromagnetic compatibility theory and applications, special radars, and unmanned aerial vehicle.

Dr. Guo is a Reviewer of the National Natural Science Foundation of China, and a Reviewer of National Science and Technology Award, a member of the Qingdao Senior Experts Association, and a Reviewer for many international journals. He has served as the local Chair for the first, second, and third International Conference on Intelligent Information Technology Applications and the 3rd IEEE International Conference on Communication and Mobile Computing.