Research Article

A Framework for Human Activity Recognition Based on WiFi CSI Signal Enhancement

Jieming Yang ⓘ**, Yanming Liu** ⓘ**, Zhiying Liu** ⓘ**, Yun Wu** ⓘ**, Tianyang Li** ⓘ**, and Yuehua Yang** ⓘ

School of Computer Science, Northeast Electric Power University, Jilin 132012, China

Correspondence should be addressed to Yanming Liu; liuyanming0515@qq.com

With the advancement of wireless technologies and sensing methodologies, many studies have shown that wireless signals can sense human behaviors. Human activity recognition using channel state information (CSI) in commercial WiFi devices plays an important role in many applications. In this paper, a framework for human activity recognition was constructed based on WiFi CSI signal enhancement. Firstly, the sensitivity of different antennas to human activity was studied. An antenna selection algorithm was proposed, which can make a choice of the antenna automatically based on their sensitivity in accordance with different activities. Secondly, two signal enhancement approaches, which can strengthen the active signals and weaken the inactive signals, were proposed to extract the active interval caused by human activity. Finally, an activity segmentation algorithm was proposed to detect the start and end time of activity. In order to verify and evaluate the methods, extensive experiments have been conducted in real indoor environments. The experimental results have demonstrated that our solutions can eliminate a large number of redundant information brought by insensitive and inactive signals. Our research results can be put into use to improve recognition accuracy significantly and decrease the cost of recognition time.

1. Introduction

Nowadays, WiFi signals cover almost every corner of people's lives, such as houses, schools, shopping malls, and buildings. If WiFi is regarded as a sensor in a sense, then WiFi-based perception systems act as the world's largest sensor network which covers all areas around us and monitors people's behaviors. With the acceleration of population aging, the demand for health monitoring is increasingly urgent, such as fall detection and health monitoring. Human activity recognition based on WiFi signals will achieve "one thing with multiple uses"; WiFi can silently perceive every action in the physical world while completing data transmission tasks. Wireless sensing technology based on WiFi signals has become an important hub linking the physical world and the information world. It has also become a research hotspot in the fields of gesture recognition [1], localization [2], and even identification [3].

In previous studies, human activity recognition systems can be categorized into four classes: wearable-based [4], vision-based [5], ambient devices-based [6], and wireless-based. Wearable sensor devices are widely used for human activity recognition especially in elder healthcare. Wearable-based human activity recognition uses hardware devices such as gyroscopes, accelerator, and barometer for recognition with high accuracy. However, these devices are expensive and inconvenience to wear. In addition, there exists limitations such as insufficient battery and people forget to wear. Vision-based methods require camera to capture human activities. However, there are still some problems such as blind spots, personal privacy, and high energy consumption. Additionally, ambient devices-based human activity recognition requires various hardware devices deployed in the environment. These ambient devices, such as pressure sensors, vibration sensors, and acoustic wave sensors are expensive, complicated to deploy, and difficult to apply in ordinary households. The movements of the human body impact the wireless signals propagation, which make it possible to capture human movements by analyzing the

received wireless signals. It has the advantages of low cost, easy deployment, wide coverage, highly penetrating [7], unaffected by light, and privacy protection.

Benefit from the widespread deployment of commercial WiFi devices in the indoor environment, using WiFi signals for human activity recognition, is a cheap solution without any additional costs [8]. In the past, some approaches based on Received Signal Strength Indicator (RSSI) had been presented for human localization [9] and human activity recognition [10]. The RSSI of wireless signals is severely affected by severe multipath and random noise in the indoor environment. Thereby, RSSI-based mechanisms have certain limitations. In recent years, new trend in device-free human activity recognition based on Channel State Information (CSI) has attracted more attention. Many previous studies have shown that CSI outperforms RSSI in human activity recognition. Therefore, in this paper, we use WiFi CSI signals for human activity recognition.

According to the background mentioned above, we have explored three issues of human activity recognition and put forward some novel proposals in this paper. The contributions of our work are summarized as follows:

(i) Based on the sensitivity of different antennas to actions, an active antenna selection approach, which makes a choice of antennas automatically, is proposed to reduce the amount of data required for subsequent calculation and analysis.

(ii) Two signal enhancement approaches were presented to achieve the enhancement of active signals. They can strengthen the interval of active signals and weaken the impact of inactive signals.

(iii) An activity segmentation algorithm was provided to detect the start and end times of activity, which can get rid of inactive signals and retain the active signal interval.

The rest of this paper is organized as follows: Section 2 reviews some related works for human activity recognition using WiFi signals. Section 3 introduces preliminaries of WiFi-CSI activity recognition. Section 4 describes the inspirations and framework. Section 5 discusses the detailed design of each module of the framework. Section 6 describes the data for experiments and presents the experimental setup. Additionally, the experimental results are presented and evaluated. Section 7 discusses the advantages and limitations of this study. Section 8 summarizes the work of this paper and looks forward to the future.

2. Related Work

Wi-Fi signals will be reflected and scattered when transmitted from the transmitter to the receiver, which causes multipath effect [11]. The overlaid multipath signals carry large amounts of information about the current features of the indoor environment. This made it possible to human activity recognition using Wi-Fi signals.

2.1. Human Activity Recognition Based on WiFi-CSI. Previous work explored the attenuation characteristics of WiFi signals [12, 13]. Coarse-grained information RSSI was used in many applications, such as environmental people counting WiCount [14], indoor localization [15], and motion tracking [11]. With the open source and release of CSI-tool, extracting CSI from commercial WiFi devices has become a reality. Due to the widespread deployment of WiFi signals, many systems based on WiFi CSI have been developed in the literature in recent years. WiFall [16] uses anomaly detection algorithms and learns specific CSI patterns to detect falls. WiFall proposed a wireless propagation model in the indoor environment under the interference of human activities and analyzed the wireless propagation model during a fall from a theoretical perspective. WiFall can realize single person fall detection with high accuracy. E-eyes [17] recognized human activity by using the moving variance of amplitude. Moving variance is more effective for those nonstationary human activities, especially those activities with sharp variations in amplitude, such as falling and jumping. However, stationary activities do not cause significant variations in amplitude in repetitive patterns, such as sleeping and sitting. In this case, the moving variance seems to be less effective. CARM [18] includes two theoretical models. One is the CSI speed model, which quantified the relationship between CSI dynamics and human movement speed, and the other is CSI activity model, which quantified the relationship between human movement speed and human activity. Guo et al. [19] combined WiFi and visual human activity recognition in HuAc. They derived the correspondence between CSI and bone-based activity recognition. In HuAc system, a mechanism of subcarrier selection was designed, which removes the first-second and the last-second data sequence of an activity according to the sensitivity of subcarriers to human activities. The HuAc system achieved the robustness of human activity recognition. CDHAR [20] is a system with WiFi-sensing radar integrated on UAVs to recognize human activities. Kernel Density Estimation (KDE) is applied in CDHAR to obtain adaptive detection thresholds and extract activity duration. CDHAR use a random subspace classifier ensemble method for classification and achieve high recognition accuracy.

Recently, deep learning methods have been widely used in human behavior recognition. Yang et al. [21] proposed a human activity recognition system with a temporal-frequency attention mechanism. In this system, a neural network model based on attention mechanism is proposed, which assigns more weight to different characteristics by imitating the human brain to focus on important information. Ding and Wang [22] proposed a WiFi CSI-based human activity recognition approach using deep recurrent neural network (HARNN), which constructs a two-level decision tree. Meanwhile, a linear regression method was also introduced to seek for the optimal parameter for the designed decision tree. Chen et al. [23] proposed a new deep learning based bidirectional long short-term memory (ABLSTM). It leverages on an attention mechanism to assign different weights for all the learned features. ABLSTM is able to achieve the best recognition performance in real

experiments. A convolutional neural network (CNN) [24] was designed to automatically extract deep features from the CSI images and achieved an average recognition accuracy of 86.3% in human activity recognition.

2.2. Antenna and Subcarrier Selection Mechanism.

Different antennas are different sensitive to static and dynamic composition in the environment. Wang et al. [18] proposed that a specific WiFi antenna link may not show significant variations in the CSI signals. Although principal component analysis (PCA) can be used to combine CSI in different subcarriers, it cannot be used to combine data from different antenna links. Therefore, they proposed three different approaches to fuse data from multiple links: majority-voting fusion, likelihood fusion, and feature fusion. In the multiple input multiple output (MIMO) system, the transceiver antennas exist in pairs; the more the amount of transceiver antennas, the higher the data dimension, which may lead to overfitting problems. To solve this problem, a subcarrier selection approach based on information theoretic learning was proposed to compensate for the overfitting problem in CSI-based localization systems [25].

2.3. Activity Segmentation Method.

Many activity segmentation algorithms were proposed in the previous work. Time-frequency analysis techniques were utilized to segment the walking movement in WiStep [26]. Activity segmentation can extract activity details and compress the data so as to improve computing speed. Wi-CR [27] took advantage of an activity indicator and a threshold to segment the activity, then counted the number of actions through a peak-finding algorithm, and determined the start and end time of each activity. WiBot [28] designed impulsive windowing approach for activity segmentation, which adopted the binary segmentation approach to detect active boundaries. WiBot allowed the start and end of gestures to be accurately identified in a continuous stream of data.

3. Preliminaries

In this section, the background knowledge of channel state information and MIMO antenna system based on CSI is summarized.

3.1. Channel State Information.

Channel state information can reflect the channel properties of communication link [29]. It describes multipath propagation of the amplitude and phase of each subcarrier in the frequency domain. Meanwhile, it contains multiple effects such as time delay, amplitude attenuation, and phase shift. CSI is more sensitive to the environment, so it can be applied to the fields such as activity recognition, gesture recognition, and motion tracking.

The wireless channel generally uses the channel impulse response (CIR) to describe the multipath effect of the channel. Under the assumption of linear time invariance, the CIR can be expressed by the following formula:

$$h(\tau) = \sum_{i=1}^{N} a_i e^{-j\theta_i} \delta(\tau - \tau_i), \tag{1}$$

where a_i represents the amplitude attenuation on the i-th path, θ_i represents the phase shift on the i-th path, τ_i represents the time delay on the i-th path, N represents the total number of propagation paths, and $\delta(\tau)$ represents the Dirichlet impulse function.

In wireless communication, the transmitted radio signals are affected by the physical environment. On the contrary, these signals can reflect changes in the physical environment. In frequency domain, multi-input-multi-output (MIMO) is modeled as

$$\mathbf{Y} = \mathbf{HX} + \mathbf{N}, \tag{2}$$

where $\mathbf{Y}$ and $\mathbf{X}$ represent the received and transmitted signal vectors, $\mathbf{N}$ represents the noise vector, and $\mathbf{H}$ represents the channel gain matrix.

CSI describes the attenuation factor of the signal on every transmission path by the channel gain matrix $\mathbf{H}$, such as signal scattering, multipath fading, power decay of distance, and other information. The multipath propagation of the signal manifests is a delay spread in the time domain, and it will cause selective fading of the signal in the frequency domain. Therefore, the channel frequency response (CFR) describes the multipath propagation of the signal using the amplitude-frequency and phase-frequency characteristics, respectively. Under the condition of unlimited bandwidth, CFR and CIR are each other's Fourier transform. The frequency response of the channel can be expressed as follows:

$$H(k) = |H(k)|e^{j\angle H(k)}, \tag{3}$$

where $H(k)$ represents the CSI of k-th subcarrier, $|H(k)|$ represents the amplitude of the k-th subcarrier, and $\angle H(k)$ represents the phase shift information.

3.2. Multiple-Input Multiple-Output Antenna System in CSI.

WiFi standards use orthogonal frequency division modulation (OFDM) in the physical layer. OFDM splits its spectrum band into multiple frequency sub-bands called subcarriers. CSI reveals a set of channel measurements depicting the amplitude and phase of every OFDM subcarrier. For example, Atheros 9590 wireless NIC generates total 56 CSI values. Intel 5300 wireless NIC reports total 30 CSI values.

CSI is extracted from the parsing packet of the Intel 5300 wireless NIC. Based on the CSI tool [30], the CSI packet received is a $N_{tx} \times N_{rx} \times 30$ matrix, where N_{tx} is the amount of transmitting antennas, N_{rx} is the amount of receiving antennas, and the third dimension is 30 subcarriers in the OFDM channel. In the commercial equipment of Intel 5300 wireless NIC, $N_{tx} = 3$ and $N_{rx} = 3$. The structure diagram of the MIMO antenna is shown in Figure 1. An antenna at the transmitter will send three data streams to the receiver. CSI packet contains 9 data streams with 30 subcarriers, which can be represented in the following format:
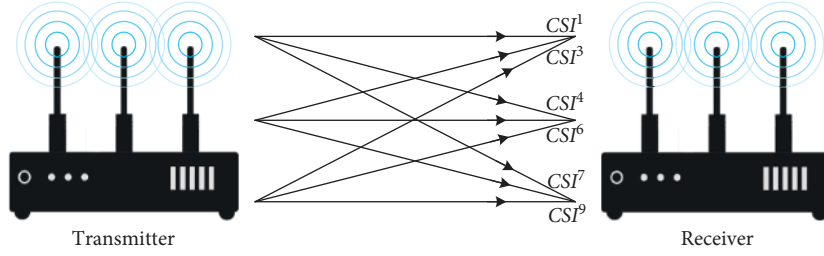
FIGURE 1: Schematic diagram of MIMO antenna structure.

$$CSI^1 = \left\{ CSI^{1,1}, CSI^{1,2}, \ldots, CSI^{1,30} \right\}$$
$$CSI^2 = \left\{ CSI^{2,1}, CSI^{2,2}, \ldots, CSI^{2,30} \right\}$$
$$\vdots \qquad\qquad\qquad\qquad\qquad (4)$$
$$CSI^8 = \left\{ CSI^{8,1}, CSI^{8,2}, \ldots, CSI^{8,30} \right\}$$
$$CSI^9 = \left\{ CSI^{9,1}, CSI^{9,2}, \ldots, CSI^{9,30} \right\}.$$

4. Framework of HAR

4.1. Inspirations

4.1.1. Antenna Selection.
Different antennas have different sensitivity to environmental perception. Thus, many works are focused on methods of subcarrier selection and fusion. Due to the diversity of the human activities and the environment, antennas are more susceptible to external factors such as the direction of human movement and the vertical dimension of the antenna, which led to the fact that antennas have different sensitivities to different actions. An antenna contains 30 subcarriers. If the antenna is not sensitive to actions, it is meaningless to select subcarriers on this insensitive antenna. Zhou et al. [31] reveal the distribution of CSI amplitude of different antennas. According to the experiments mentioned above, different antennas have different sensitivity to the same activity. For example, in the bend movement, one antenna is insensitive, while the others are sensitive. Based on the above inspiration, we have explored the relationship between antennas and proposed an antenna selection mechanism to remove those antennas that are not sensitive to the activity.

4.1.2. Enhancement of Activity Signal.
In previous work, filter, outlier elimination, and interpolation are often used for data preprocessing, such as Butterworth filter [32], Kalman filter [33], Hampel filter, and discrete wavelet transform (DWT) [34]. However, these methods only reduced the noise instead of enhancing activity signals. If the difference between the active signal and the inactive signal can be augmented, the active signals will be enhanced and the inactive signals will be weakened. Based on the above inspiration, a signal enhancement approach is proposed. The enhanced signals will clearly indicate the active intervals; meanwhile, those inactive ones will be further weakened, which will suffice to separate the active signals and the inactive ones.

4.1.3. Activity Segmentation of Start and End Times.
In the entire CSI sequence, the signals caused by human activity account for only a small part. Most of the signals are composed of inactive signals before and after the action. If the features of the entire CSI sequence are extracted and input into the classifier training, a large number of inactive signals will increase the amount of calculation and affect the accuracy. In the previous work, the variance of the phase difference between the antennas is used to detect a fall [35, 36]. Hilbert transform extracts multiple envelopes to achieve activity segmentation. In our paper, an activity segmentation algorithm is proposed to detect the start and end times of activities based on signal enhancement.

4.2. Framework of HAR.
The HAR framework consists of the antenna selection module, signal enhancement module and activity segmentation module in Figure 2. We describe the details of every module in Section 5.

The antenna selection module selects the antenna which is sensitive to different activities and abandons the others. The signal enhancement module includes SavitzkyGolay filter and interpolation and signal enhancement. Among these studies, this paper focuses on signal enhancement. Two approaches have been proposed for signal enhancement, N-iteration signal enhancement (NISE) and P-signal enhancement (PSE). The signal enhancement amplifies the signal which implies activity and weakens the signal which indicates inactivity. The activity segmentation module segments the active and inactive parts of the signal. In this module, an activity segmentation algorithm is proposed, which aims at detecting the intervals of the activity.

5. Construction of HAR Framework

5.1. Antenna Selection Module.
In this section, a MIMO antenna system which consists of one transmitter and three receivers comes into use. The raw signals of various activities were analyzed based on a large number of experiments as shown in Figure 3. The results show that the existence of insensitive antennas is inevitable rather than accidental. Three representative activities were selected, such as vigorous movement (bend), slight movement (clap), and continuous repetitive stable movement (walk).

It can be seen from Figure 3 that there exists an antenna which is not sensitive to human activity in the 1×3 antenna system. Thus, it is named insensitive antenna. The signal on the insensitive antenna is seriously interfered by noise and
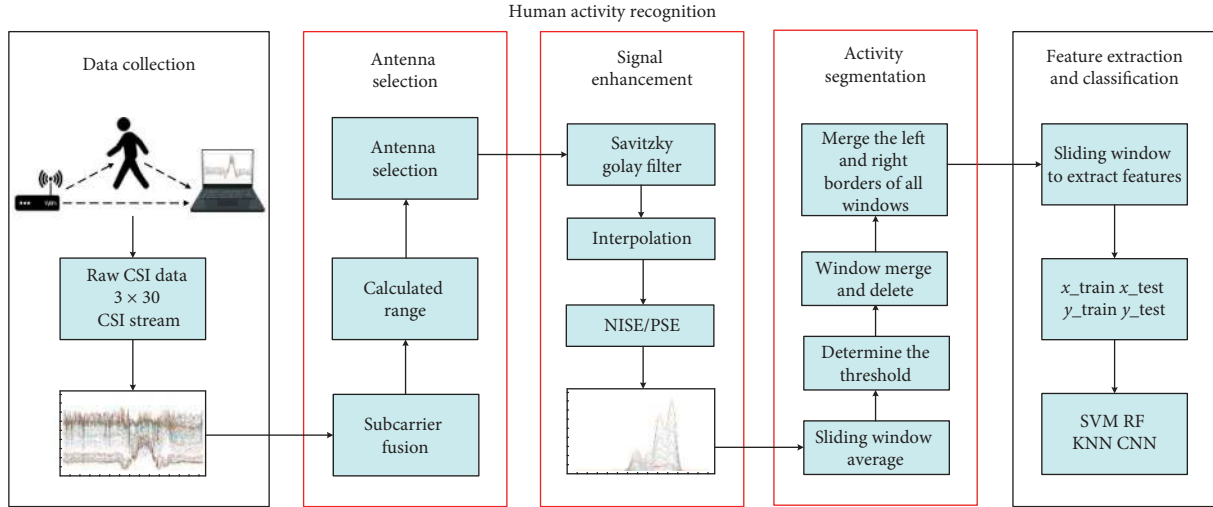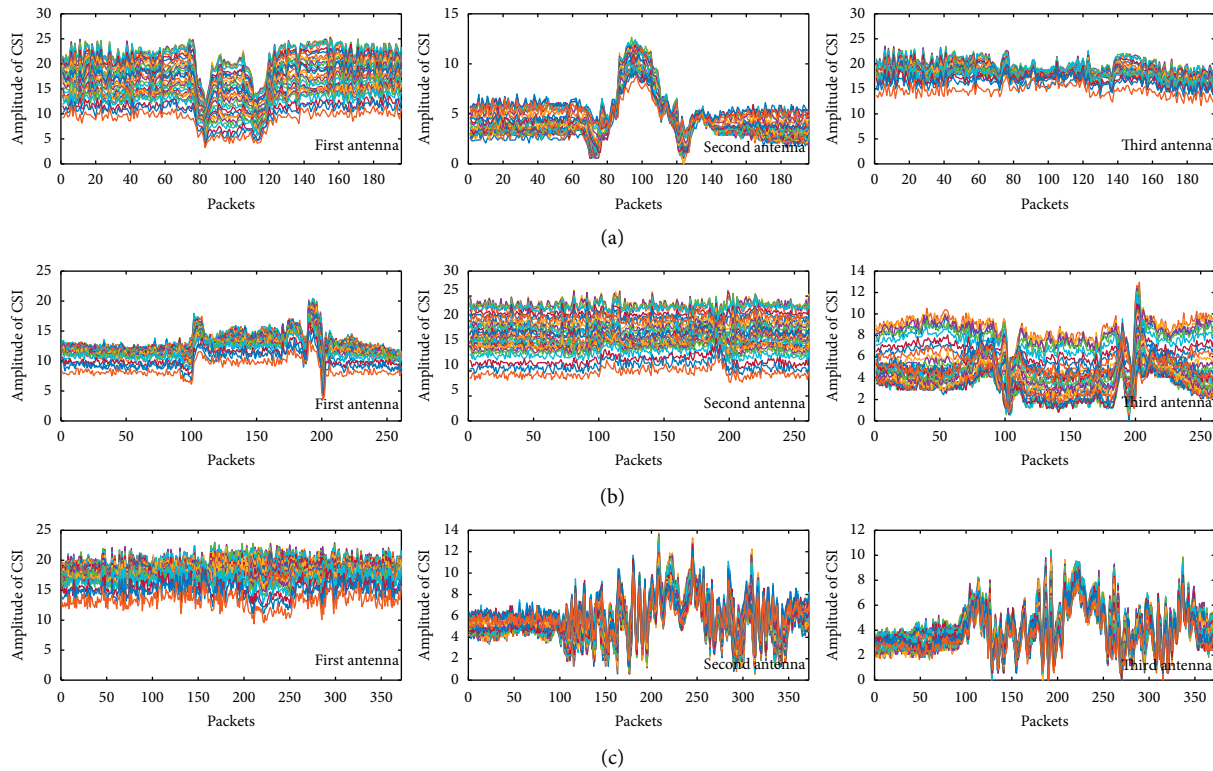
FIGURE 2: The framework of HAR.

FIGURE 3: Raw signals of three antennas with different activities: (a) bend; (b) clap; (c) walk.

hardly reflects the human activity. If this antenna is used in the final classification, the recognition accuracy will be seriously degraded. If the insensitive antenna will be abandoned, the characteristic information would not lost due to the information redundancy and correlation among these antennas.

It can be found that the sensitivity of different antennas to activity is different. Insensitive antenna contains a significant characteristic that the amplitude of CSI is relatively stable, whereas the signal of the sensitive antenna will change

obviously. The reason for the existence of insensitive antennas may be related to factors such as the experimental environment, physical antenna placement, and human body orientation. Our purpose is to find and remove insensitive antennas without considering the quantitative relationship between the antenna and the above influencing factors.

Based on the above research, we propose an adaptive antenna selection approach, which choose or reject the antennas according to the sensitivity of different activities. The experiments make a comparison to 30 subcarriers

between the insensitive antenna and the sensitive antenna. The results revealed that the signal change trend and the activity range of the sensitive antenna are consistent. We calculate the average of 30 subcarriers and form a data sequence, and the activity interval of the sensitive antenna is very obvious, such as the first and second antennas as shown in Figure 4(a). Meanwhile, the insensitive antennas, such as the third antenna, are stable with a small range of fluctuations and insensitive to human activities.

In order to further distinguish the sensitive to human activities of the antennas, the sliding window variance approach was adopted in the analysis of the three CSI streams. As shown in Figure 4(b), the first antenna is the most sensitive to movement, and the third antenna is the least sensitive. It means that the difference between them is expanded significantly. Finally, it can conclude that the first antenna, which is the most sensitive to human activities, is the best choice. The antenna selection algorithm is described as follows (Algorithm 1).

5.2. Signal Enhancement Module

5.2.1. Stability Measurement Based on Variance Theory. In the theory of probability and statistics, variance is a measure of the dispersion degree of a set of data, which is used to describe the distance between the sample and its mean center. The CSI measurement of a subcarrier is denoted as $A = \{A_1, A_2, \ldots A_i \ldots A_n\}$, and the difference between the measured value A and the true value $\widehat{A}$ is denoted as $\delta = \{\delta_1, \delta_2 \ldots \delta_i \ldots \delta_n\}$, where n represents the number of samples, $\delta_i = A_i - \widehat{A}$. The variance can be defined as

$$S^2 = \frac{1}{n} \sum_{i=1}^{n} \delta_i^2 = \frac{1}{n} \sum_{i=1}^{n} \left(A_i - \widehat{A}\right)^2. \qquad (5)$$

However, the true value $\widehat{A}$ is unknowable and δ_i cannot be obtained; so, formula (5) has only theoretical significance. In practical applications, the arithmetic mean $\overline{A}$ can often be used to represent the true value $\widehat{A}$. V_i can be defined as $V_i = A_i - \overline{A}$. δ_i and V_i have the following mathematical relationship [37]:

$$\sum_{i=1}^{n} \delta_i^2 = \frac{n}{n-1} \sum_{i=1}^{n} V_i^2. \qquad (6)$$

The S^2 can be modified as follows:

$$S^2 = \frac{1}{n} \sum_{i=1}^{n} \delta_i^2 = \frac{1}{n} \frac{n}{n-1} \sum_{i=1}^{n} V_i^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(A_i - \overline{A}\right)^2, \quad (7)$$

where A_i represents the amplitude of sample i, $\overline{A}$ represents the mean center of the sample, and n represents the number of samples. In the CSI signal, the signal time series reflects the change of human activity with time. If the variance is calculated in the entire time series, it will be meaningless and only reflects the average stability of the entire process. In a local range, the variance can represent the discrete degree of instantaneous activity. If the signal in the inactive range tends to be stable, the variance in the sliding window is small,

and if the signal in the active range is unstable, the variance in the sliding window will grow larger. Based on the above ideas, this paper introduces a sliding window to calculate the variance of the local range to measure the stability and instability and then roughly distinguishes between active signals and inactive signals.

5.2.2. N-Iterations Signal Enhancement (NISE). The raw CSI signal contains a lot of noise; the key activity signal range is submerged in the noise. Most of the previous work was based on filters to remove noise interference and rarely considered enhancement of active signals and suppression of inactive signals. Based on the above inspiration, formula (7) is used to describe the stability of the data in samples. This approach can strengthen the activity signals, but the enhanced signal has obscure activity boundaries as shown in Figure 5(a).

To solve the above problem, we proposed a signal enhancement approach based on *N*-iterations (where *N* is the number of iterations), which means that the signal was enhanced multiple times with the same approach. As shown in Figure 5(b), in a single subcarrier, N-iterations signal enhancement (NISE) outperforms the above approach.

The iterative structure with the sliding window is shown in Figure 6. The CSI amplitudes of every subcarrier is denoted as $S = \{1, 2, 3, \ldots, n\}$, where n represents the number of packets. We calculated the variance of the raw signal in the sliding window. These variances form a new variance sequence $A = \{A(1), A(2), A(3) \ldots A(k)\}$, where $k = n - Wz + S$, Wz is the size of slide window, and S is the step, which is used to calculate the variance of the next round and achieve $B = \{B(1), B(2), B(3) \ldots\}$. It should be noted that the size of data sequence will changes for each iteration.

The NISE enhanced signals of the three antennas is shown in Figure 7. The signals in the left picture are the raw signals, it can be seen that the active and inactive parts of the raw signals are difficult to distinguish, and their boundaries are blurred. On the right are the enhanced signals. NISE enhances the active signal and weakens the inactive signal, which leads to the fact that the active window boundary of the enhanced signal is clear. By means of signal enhancement, the fact that there exists insensitive antenna is affirmed in the experiment. Sensitive antenna (a) and (b) have overlapping active signal windows, whereas the insensitive (c) did not show the same characteristic after being strengthened, and its windows are still scattered.

The pseudocode of the NISE algorithm is given as follows (Algorithm 2).

5.2.3. P Signal Enhancement (PSE). Considering the issue that NISE requires multiple rounds of iterative calculations with high computational overhead, a P-signal enhancement (PSE) is proposed in the following studies. The formula of P-signal enhancement is defined as follows:

$$S = \sum_{i=1}^{n} \left(A_i - \overline{A}\right)^{2p}, \quad p = 1, 2, 3 \ldots, \qquad (8)$$
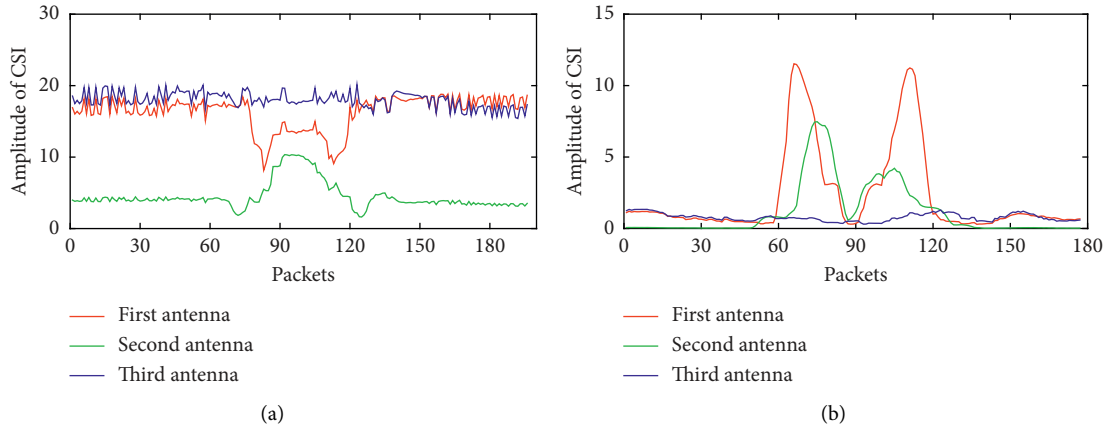
(a)

(b)

FIGURE 4: Comparison of sensitivity of different antennas. CSI sequences (a) formed by fusion of 30 subcarriers of each antenna and (b) after sliding window variance based on (a).

Input: S—the sequential data of CSI which contain 3 antennas with 30 subcarriers
W—the size of sliding window
$step$—the step size of window movement
Output: S_j—the sequential data of the antenna which are most sensitive to activity
Step 1: for each antenna S_j in S
Step 2: calculate the mean sequential data (F_j) of 30 subcarriers in S_j
Step 3: N_j = the length of F_j
Step 4: $E_j = \emptyset$
Step 5: for (int $k = 0$; $k + W \leq N_j$; $k = k + step$)
Step 6: calculate the variance (v_k) of sequential data in sliding window
Step 7: append v_k to E_j
Step 8: end
Step 9: $R_j = \max (E_j) - \min (E_j)$
Step 10: end
Step 11: return antenna S_j whose corresponding R_j is maximum

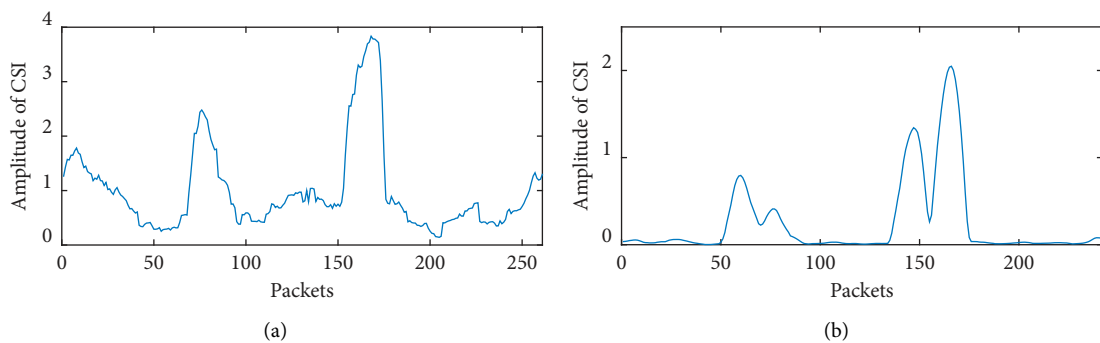ALGORITHM 1: Antenna selection.

(a)

(b)

FIGURE 5: Sliding window variance and N-iterative signal enhancement. (a) The sliding window variance of subcarrier No. 24 of bending action; (b) N-iterative signal enhancement of No. 24 subcarrier of bending action.

where A_i represents the amplitude of sample i, $\overline{A}$ represents the mean center of the sample, and p represents a natural number. $2p$ ensures that the distance between the sample and its mean is positive, because we focus on measuring the degree of deviation, negative distance is meaningless. The raw CSI signals of different qualities have different p values.

If the CSI signals are less interfered by environmental noises, then a smaller p value can obtain inspiring effect brought by means of signal enhancement. Here, the value of p is 2; the value of p will be discussed in Section 6.

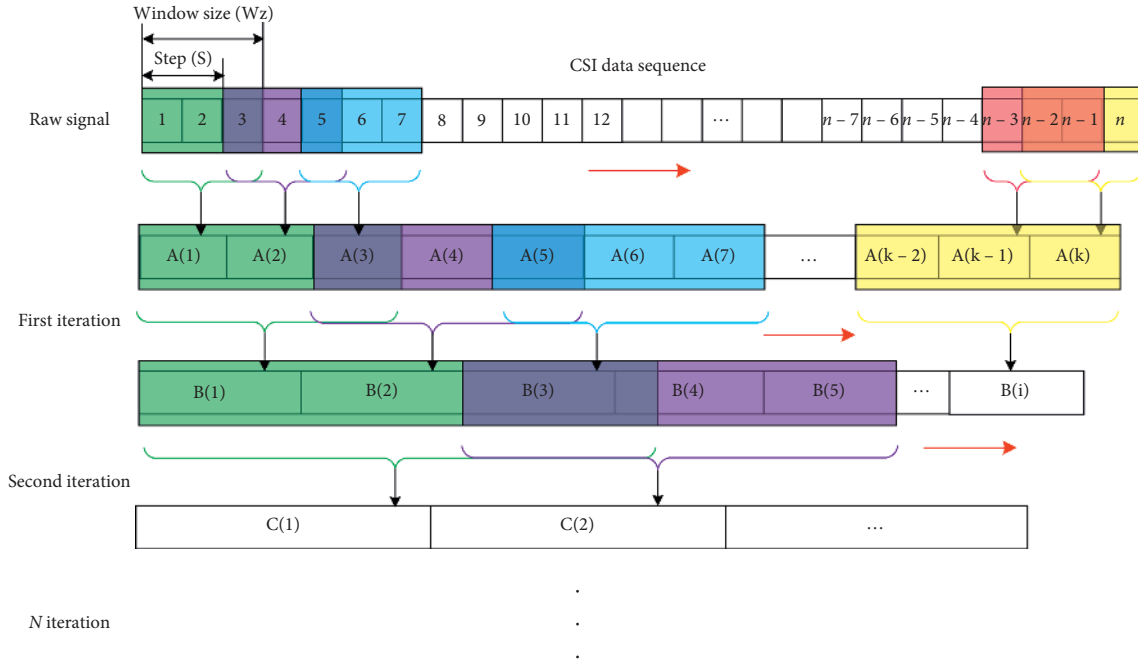The pseudocode of PSE algorithm is given as follows (Algorithm 3).

Figure 6: The iterative structure of the sliding window.

The PSE enhanced the signals of the first and second antennas in the bend activity, which is shown in Figure 8. It is not difficult to conclude that PSE is comparable to NISE.

5.3. Activity Segmentation Module.

Activity segmentation aims to detect the start and end times of activity. Figure 9(a) has shown the segmentation of a single subcarrier. In our experiments, 30 subcarriers on each antenna were adopted to explore active intervals. Figure 9(b) has shown the result obtained by combining the active intervals of all subcarriers, which describe the start and end times of human activity. In a word, the method, which is used to segment a single subcarrier, is evolved to deal with all subcarriers at one time and form a comprehensive segmentation of human activity. Therefore, the activity segmentation algorithm for integrating the activity interval of all subcarriers is proposed. The pseudocode of the activity segmentation algorithm is given as follows (Algorithm 4).

5.4. Feature Extraction.

Plenty of relative studies summarize the feature extraction methods of WiFi-CSI human activity recognition. We extract the following features from CSI amplitude to classify human activity: (1) mean, (2) normalized standard deviation (STD), (3) maximum and minimum, (4) skewness and kurtosis, (5) median absolute deviation (MAD), (6) signal range, (7) interquartile range (IR), (8) signal entropy, and (9) velocity of signal change. These features, which extracted from CSI amplitude, are all set to be the input of the classifier.

5.5. Classification.

Various classification methods have been applied to classify human activities. In order to discuss whether or not our approaches mentioned above can achieve better performance and validity, machine learning methods and deep learning methods were applied to verify the effectiveness of the proposed approaches.

Machine learning classifiers such as support vector machine (SVM), random forest (RF), and K nearest neighbor (KNN) were applied in our experiments. SVM is a supervised learning model in machine learning, which is used to analyze data and recognize patterns. In order to solve the nonlinear classification problem, a kernel function is used to map input samples into a high-dimensional feature space. It can find the maximum margin hyperplane in the transformed feature space. Random forest (RF) is based on ensemble learning methods for classification and regression. The RF classifier consists of a collection of single decision trees, each of which is grown by randomly drawing samples and replacing them. RF improves the classification performance of a single-tree classifier by constructing decision trees with random methods, such as the bootstrap (bagging) method. The random forest selects the tree with the most votes to classify it in all the forests. KNN is a basic classification and regression method, which is an optimization problem of finding the closest point in a scale space. KNN classifies by measuring the distance between different feature values. This distance is determined by Euclidean distance or Manhattan distance.

Convolutional neural network (CNN) is a kind of feedforward neural network with convolution operation and deep structure and, therefore, is regarded as one of the
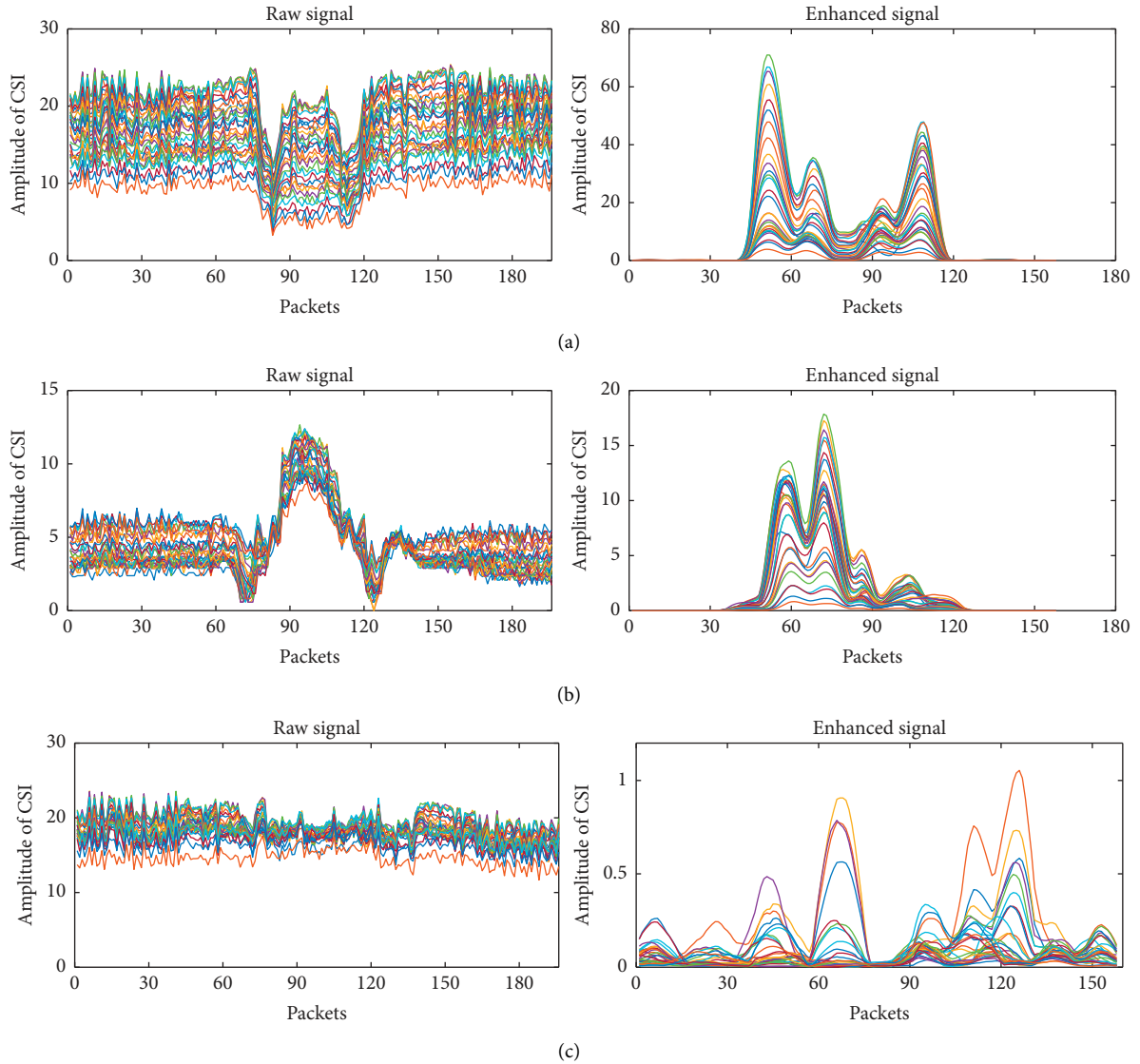
Figure 7: Three antennas raw signal and enhanced signal of bend movement based on NISE. (a) First antenna; (b) second antenna; (c) third antenna.

representative algorithms of deep learning. It has the ability of representation learning and classifying input information according to its hierarchical structure.

6. Implementation and Evaluation

6.1. Experimental Setup. The experimental environment is built on off-the-shelf devices. The experimental data acquisition system consists of two devices. Two ThinkPad X200 laptops equipped with an Intel 5300 NICs served as the transmitter and receiver, each of which has three external 4 dBi Gain omnidirectional antennas. The laptop is installed Ubuntu 12.05 with a modified Intel NIC driver and the kernel version is 4.2.0. In order to prevent interference of many devices working at 2.4 GHz, the experimental system is designed to support two frequency bands 2.4 GHZ and 5.2 GHZ. The software used in our experiments is the open-source CSI-tools presented by Shangguan et al. [38]. Python software was used to analyze the collected data as described in the methodology section and MATLAB software was used to achieve the visualization of results. The experimental hardware is shown in Figure 10.

These experiments were carried out on three typical indoor environments with different layout schemes. The experimental scenarios are shown in Figure 11. Three volunteers, of height between 165 cm and 185 cm, join these experiments. Each volunteer performs specific activities individually. The distance between the transmitter and receiver is 2.5 m–3m, and the vertical height is 1.2 m. The experiments were implemented on IEEE 802.11n monitor mode at 5.2 G WiFi frequency in order to get rid of the crowded 2.4 GHz interference in the experimental environment. The sampling rate is 30 packets per second.

Input: S_{ij}—the sequential data of i-th subcarrier and j-th antenna for CSI signal
W—the size of sliding window
P—the number of iterations
$step$—the step size of window movement
Output: SE_{ij}—the enhanced sequential data of i-th subcarrier and j-th antenna
Step 1: $S = S_{ij}$
Step 2: for $(m = 0; m < P; m++)$
Step 3: $N =$ the length of S
Step 4: for $(k = 0; k + W \leq N; k = k + step)$
Step 5: $ST = \varnothing$
Step 6: calculate the variance (v_k) of sequential data in sliding window from S
Step 7: append v_k to ST
Step 8: end
Step 9: $S = ST$
Step 10: end
Step 11: $SE_{ij} = S$
Step 12: return SE_{ij}

ALGORITHM 2: N-iteration signal enhancement.

Input: S_{ij}—the sequential data of i-th subcarrier and j-th antenna for CSI signal
W—the size of sliding window
$step$—the step size of window movement
Output: SE_{ij}—the enhanced sequential data of i-th subcarrier and j-th antenna
Step 1: $N =$ the length of S_{ij}
Step 2: for $(k = 0; k + W \leq N; k = k + step)$
Step 3: $ST = \varnothing$
Step 4: calculate formula (6) (v_k) based on sequential data in sliding window from S_{ij}
Step 5: append v_k to ST
Step 6: end
Step 7: $SE_{ij} = ST$
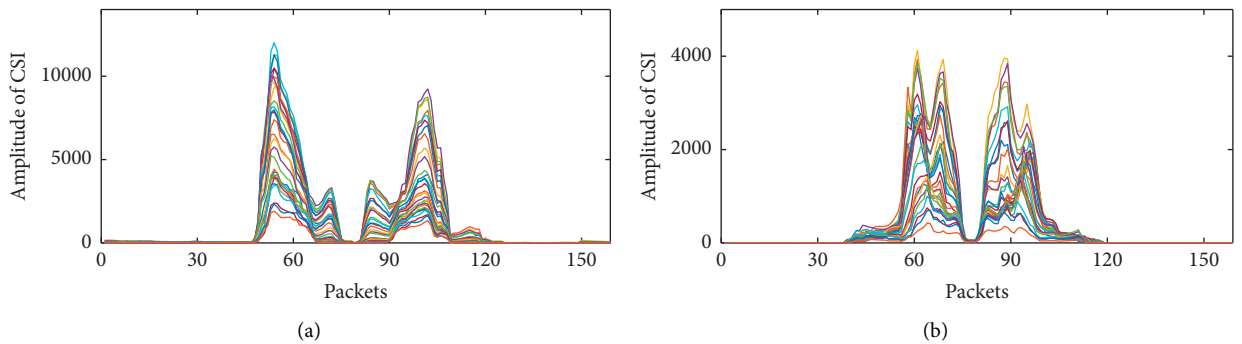Step 8: return SE_{ij}

ALGORITHM 3: P-signal enhancement.

(a)

(b)

FIGURE 8: Enhanced signal based on PSE. (a) First antenna; (b) second antenna.

6.2. Dataset Description.

Three volunteers were recruited to perform eight daily activities including bend, call, clap, drink, sit, squat, walk, and wave in three different scenarios. Each volunteer was required to finish the activities individually for a period of 5–20 seconds. It is important to note that the volunteer remains stationary in addition to perform specific activities. In order to simulate the activities under real conditions, items on the table will be moved randomly. During the experiment, the door of the room remains closed and there was no furniture to move. In addition, both the transmitter and receiver are placed in the line-of-sight (LOS) conditions.

Six-hundred data files of all activity from three volunteers were collected. Datasets are described in Table 1. A sliding window was used to extract features from samples to
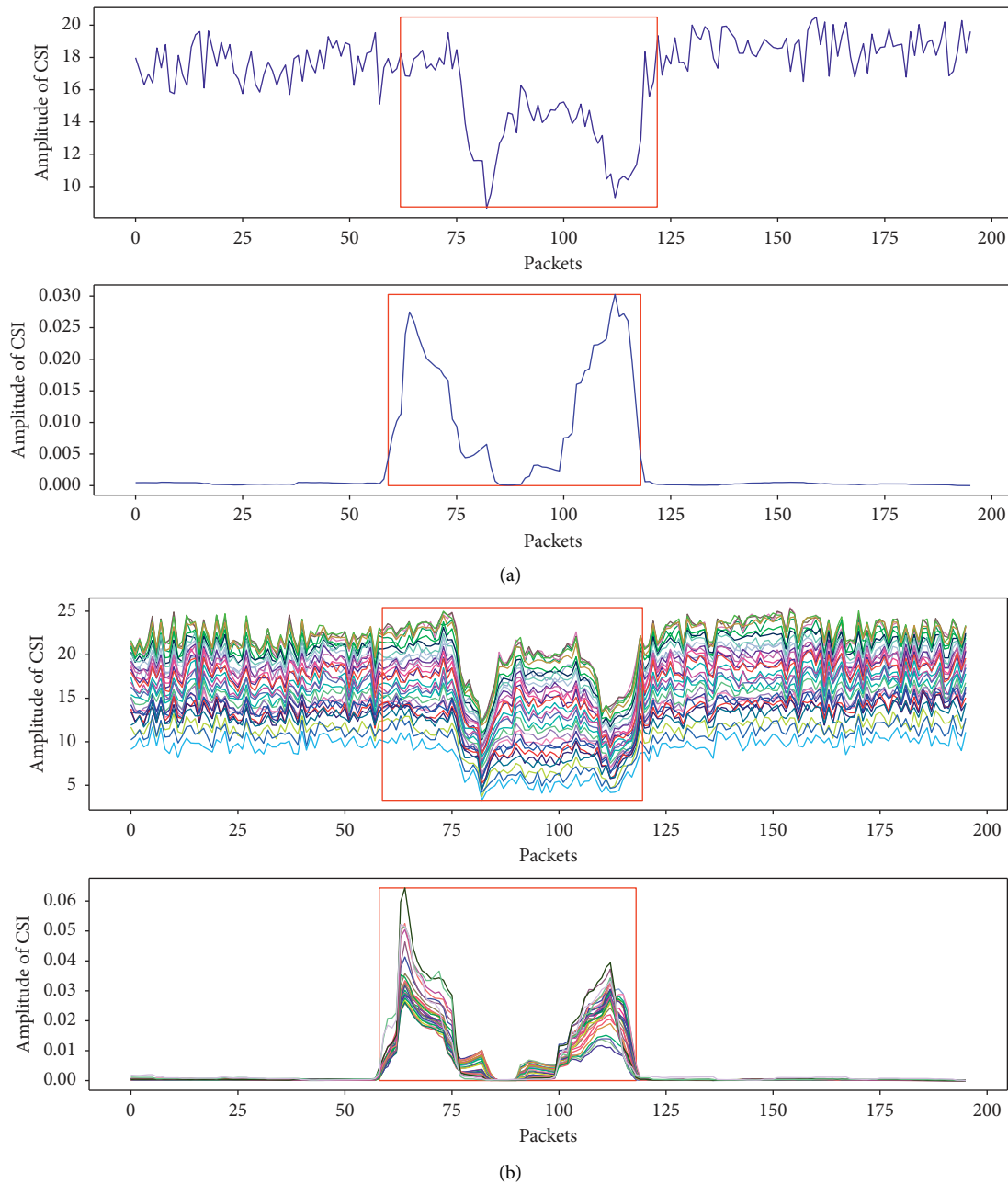
FIGURE 9: The raw signal and segmented signal of (a) a single subcarrier and (b) all subcarriers.

generate labeled feature data. The data set was divided into 90% training and 10% testing to build three classifiers, and we also measured the five-cross validation accuracy.

6.3. Performance of Human Activity Recognition. This section discusses the impact on human activity recognition from the following three aspects.

6.3.1. Impact of the p Value in PSE. The effect brought by the p value was observed on the signals. As shown in Figure 12, the signal becomes sharp and the range of activity tends to be constant with the increase of p value. In order to obtain the start

and end times of the activity, this paper maps this time range to the raw signal and segments the activity. Therefore, we pay more attention to the boundary of the enhanced signal rather than the amplitude shape. The performance results under different p values are shown in Figure 13. It can be seen that the performance is best when p value is 2, and the accuracy and precision are 96.86% and 97.81%, respectively. With the increase of p value, the system performance did not continue to improve.

6.3.2. Impact of Sliding Window Size in NISE and PSE. The size of the sliding window is the key to signal enhancement. According to our researches, the appropriate size of the sliding window is closely related to the sampling frequency

Input: S_e—the enhanced signal of CSI
W—the size of sliding window
N—the length of the sequential data of CSI
$step$—the step size of window movement
Output: T_s—the start time point of activity
T_e—the end time point of activity
Step 1: for each subcarrier S_j
Step 2: $V = \emptyset$
Step 3: for $(k = 0; k + W \leq N; k = k + step)$
Step 4: calculate the mean(m_k) of sequential data in sliding window from S_j in S_e
Step 5: append m_k to V
Step 6: end
Step 7: $V_S = V$
Step 8: sort V_S in ascending order
Step 9: t = the numerical value of third quartile (75%) in sorted V_S
Step 10: filter out the value that is less than t in V
Step 11: the range of the remaining continuous data in V is the start time(v_{js}) to the end time(v_{je}) in sequential data for the activity
Step 12: end
Step 13: $T_s = \min(v_{js})$
Step 14: $T_e = \max(v_{je})$
Step 15: return T_s, T_e

ALGORITHM 4: Activity segmentation.

FIGURE 10: Experimental hardware.

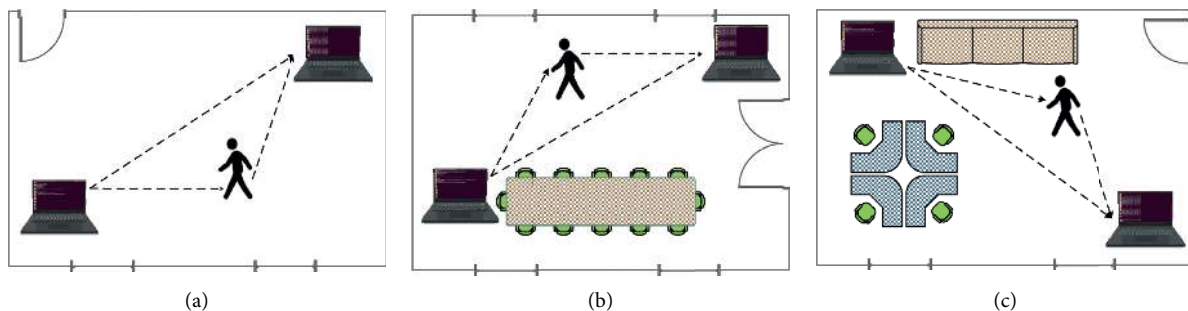(a)   (b)   (c)

FIGURE 11: The scenarios of human activity experimental: (a) empty room, (b) meeting room, and (c) research room.

(our sampling frequency is 30 Hz). The amount of data in sliding window is related to the duration of human activity and reflects the transient movement. If the sliding window is too small, human activities will be oversegmented and cannot contain integral human behavior. Conversely, if the sliding window is too large, it cannot reflect this microvariation of human behavior and only indicate the overall changes. The relationship between the size of the sliding window and the signal enhancement performance is shown in Figure 14. It can be found that 20 is the best size of the sliding window. As the

sliding window increases, the system performance decreases rapidly. According to a large number of experiments, it can be concluded that the signal enhancement performance is the best when $W = F/1.5$, where the sliding window size is denoted as W and the sampling frequency is described as F.

6.3.3. Impact of the Experimental Scenarios. We evaluated the performance of human activity recognition in three experimental scenarios. Table 2 makes a comparison

TABLE 1: The details of datasets in three scenarios.

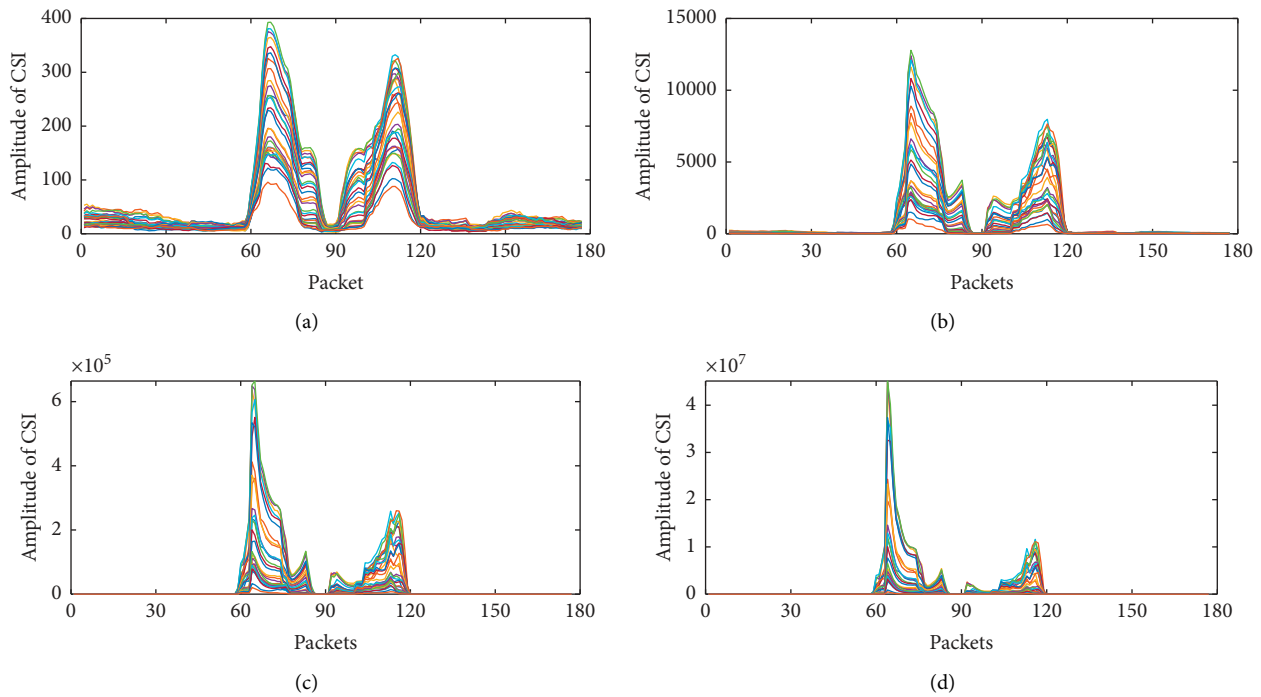| Scenario	Volunteer	Number of data files	Number of activities per file	Time per file (s)
Empty room	A	600	1	8
	B	600	1	8
	C	600	1	8
Meeting room	A	200	3	15
	B	200	3	15
	C	200	3	15
Research room	A	150	4	20
	B	150	4	20
	C	150	4	20

FIGURE 12: The impact of different p values on signal enhancement. (a) $p = 1$, (b) $p = 2$, (c) $p = 3$, and (d) $p = 4$.
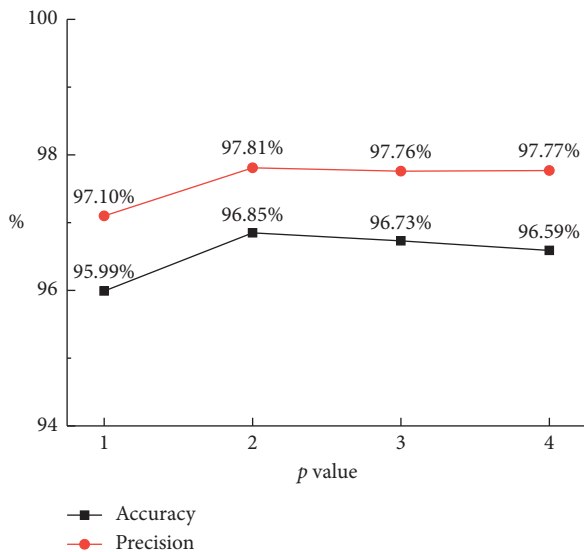

FIGURE 13: Performance results under different p values.

between different activities on different experimental scenarios. The overall performance in the meeting room is better than the other two experimental scenarios. It can be seen from Table 2 that the average accuracy of walk is the highest, because it is a continuous repetitive action with a single action pattern. The difference between individuals is relatively insignificant. The average recognition accuracy of clap and drink is a little poor, because these actions are often accompanied by other body movements at the same time. These movements are complex and diverse and have no fixed pattern, which makes it is difficult to recognize.

6.4. Evaluation

6.4.1. Comparison between Different Approaches. The diversity of individual human activities determines the diversity of CSI information, which means that different persons possess different movement patterns (such as posture, speed, range of mentioned, and height). In order to verify the performance of these methods proposed in this

TABLE 2: Accuracy of different activities in each experimental scenario.

Scenarios	Approach	Bend	Call	Clap	Drink	Sit	Squat	Walk	Wave
	Raw signal	93.63	96.46	88.38	88.81	93.78	89.74	98.76	87.49
Empty room	NISE	96.94	99.46	92.09	92.42	98.63	90.58	100	97.74
	PSE	96.63	98.98	91.33	91.48	97.83	91.20	100	96.04
	Raw signal	90.09	89.37	91.57	91.26	92.56	94.49	93.90	92.76
Meeting room	NISE	95.07	95.37	95.02	98.30	96.39	98.59	100	99.35
	PSE	92.91	95.17	92.75	97.96	98.51	99.20	100	99.49
	Raw signal	81.08	82.18	80.52	80.58	81.78	82.88	86.59	83.35
Research room	NISE	96.30	91.60	94.37	92.38	99.10	97.68	99.10	89.01
	PSE	94.28	90.77	91.84	92.22	98.84	99.11	98.96	91.10
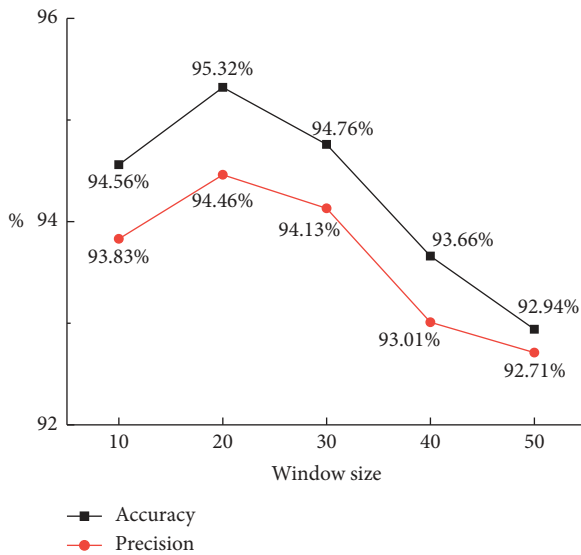
FIGURE 14: The impact of sliding window size on performance.

FIGURE 15: Performance of different approaches on three volunteers and their fusion data using RF classifier in an empty environment.

paper, machine learning and deep learning methods were applied to our system. Three volunteers A, B, and C were recruited to take part in the experiment. Meanwhile, any combination of these datasets is utilized and described as A-B, A-C, B-C, and A-B-C. The performance of different approaches on three volunteers and their fusion data using RF classifier in an empty environment is shown in Figure 15.

It can be shown in Figure 15 that the performance of NISE and PSE is significantly better than the raw signals. Among them, the performance of NISE is slightly better than PSE. The average recognition accuracy of A and B is better than C. According to our observation, volunteers A and B are both male and have similar activity styles. Volunteer B who exercises regularly and obtains 96.82% of average recognition accuracy. Volunteer C is a female who rarely exercises and does not have standard activity. Volunteers B and C are one male and one female, respectively; therefore, there are differences in posture and height. Based on above reasons, it can be confirmed that the greater similarity of the height, posture, and activity styles between volunteers, the better the recognition performance of the system.

The CNN was constructed with two convolutional layers and two pooling layers. The size of the convolution kernel is 5×5 and the size of the pool is 2×2. The number of iterations epoch in CNN is set to 200. The data extracted from
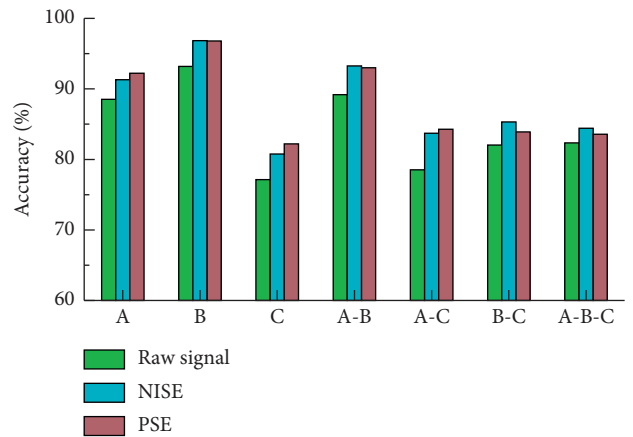
CSI sequence generated a matrix, the rows of which correspond to subcarriers and the number of columns of which is equal to the size of the slide window. The matrix with 30×60 is used as input data of CNN in our experiments. With three volunteers and their fusion data achieved in an empty environment, the performance of CNN based on different approaches is shown in Figure 16.

Experiment results show that the performance of CNN based on NISE and PSE is better than that based on raw signals; therefore, it is not difficult to conclude that the enhanced and segmented signal can obtain better recognition accuracy than the raw signal in deep learning. NISE and PSE can obtain 93.81% average of recognition accuracy in A-B-C datasets. Moreover, it is worth while to note that the performance of NISE and PSE in deep learning is stable in the fusion datasets.

Moreover, the confusion matrices were built to evaluate our system. Figure 17 shows the confusion matrix of the experiment results created by NISE and PSE in the RF classifier. Each row represents an actual class, where each column represents a predefined class. The average accuracy is 95.75% in NISE and 94.5% in PSE.

6.4.2. Comparison between Different Classifiers. We compared three classifiers with different architectures (SVM, RF, and KNN); SVM implemented 10-fold cross
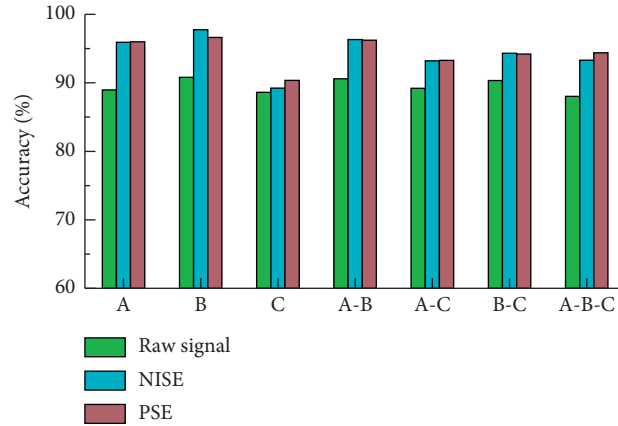
FIGURE 16: Performance of CNN based on different approaches on three volunteers and their fusion data in an empty environment.

FIGURE 17: The confusion matrices created by (a) NISE and (b) PSE.

validation. The number of trees is 200 in RF classifier, and the value of k in KNN classifier is 13, which provides higher classification accuracy and avoids obvious overfitting. Based on these results in Figure 18, we can draw a conclusion that KNN has the best recognition performance, Volunteer B's actions are the most standard ones, and the recognition accuracy of the three classifiers is almost equal.

6.4.3. Comparison between Different Indicators. To evaluate the performance of the proposed approaches, accuracy, precision, recall, and F1 score were used to analyze results of the experiments. The RF classifier was used in volunteer B dataset to analyze classification indicators; the formulas of the precision and recall for the category C_i are defined as

$$P_i = \frac{TP_i}{TP_i + FP_i},$$

$$R_i = \frac{TP_i}{TP_i + FN_i}, \tag{9}$$

where TP_i is the number of the activities that is correctly classified to category C_i, TN_i is the number of the activities that is correctly classified to other categories excluding the category C_i, FP_i is the number of the activities that is misclassified to the category C_i, and FN_i is the number of the activities belonging to category C_i, which are misclassified to other categories. To evaluate performance average across categories, the microaveraging and macroaveraging were used in our experiments. Microaveraging is obtained by summing the over all individual decisions. Macroaveraging is evaluated "locally" for each
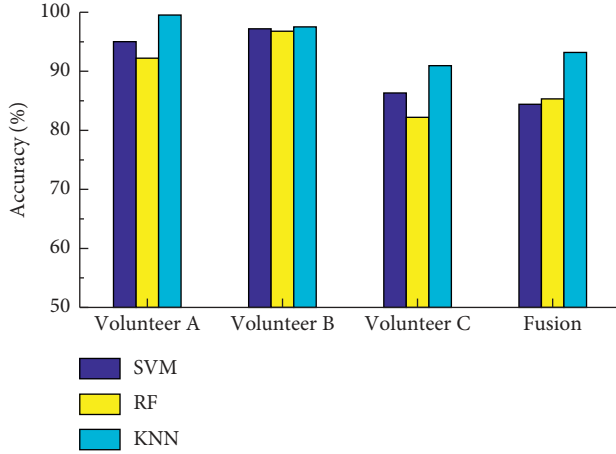
Figure 18: Comparison between different classifiers using NISE approach.

Figure 19: Comparison of the classifier indicators in the three approaches using RF classifier in an empty room.

category and then "globally" by averaging over the results of the different. The microprecision, microrecall, macroprecision, and macrorecall may be obtained as

$$P_{\text{micro}} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{\sum_{i=1}^{|C|} \text{TP}_i}{\sum_{i=1}^{|C|} (\text{TP}_i + \text{FP}_i)},$$

$$R_{\text{micro}} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{\sum_{i=1}^{|C|} \text{TP}_i}{\sum_{i=1}^{|C|} (\text{TP}_i + \text{FN}_i)},$$

$$P_{\text{macro}} = \frac{\sum_{i=1}^{|C|} P_i}{|C|},$$

$$R_{\text{macro}} = \frac{\sum_{i=1}^{|C|} R_i}{|C|}.$$

(10)

In our evaluation, "macro" is used to analysis recall and precision and "micro" is adopted for F1. The micro-F1 and accuracy are defined as follows:

$$\text{F1}_{\text{micro}} = \frac{2 \times P_{\text{micro}} \times R_{\text{micro}}}{P_{\text{micro}} + R_{\text{micro}}},$$

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}},$$

(11)

where TP = true positives, FP = false positives, TN = true negatives, and FN = false negatives. Among these indicators, accuracy and precision are the most important measures in our studies. Accuracy indicates the proportion of correct recognition in all activities. Precision can identify the proportion of human activity in all detected activities, so it is a measure of false alarms. The recall rate provides the proportion of activities that the system correctly recognizes in actual activities. F1 score is the harmonic mean of these two metrics. It can be seen from Figure 19 that NISE and PSE are significantly better than the raw signal. Among the classifier indicators, precision performed best, reaching 97.8%, which also shows that our system
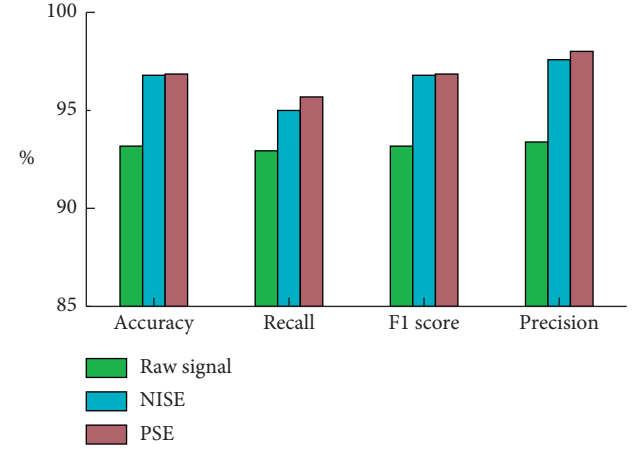
has a low false alarm rate. Accuracy reflects the overall performance of the system, reaching 96.82%.

7. Discussion

In fact, only a small portion of the whole signal is available to represent the characteristics of human activities. All the signals are used to train the classifier, which will lead to large amount of calculation, long recognition time, and low efficiency. Thus, this paper puts forward to remove insensitive antenna signal from the raw signal and construct a framework, which consists of the antenna selection module, signal enhancement module, and activity segmentation module. The solutions described above are implemented in our system, which improves the recognition accuracy and reduces the time required for recognition. The experimental environment in this paper is closer to actual needs, such as fall detection, home safety detection, and other scenarios that require real-time detection. The methods proposed in this paper aims to achieve real-time detection and recognition and will be widely used in real-life scenarios.

However, there are still many limitations in our work. First, only the CSI amplitude information is used in current studies, and more accurate CSI phase information will be still needed in the future. Secondly, existing data collection equipment requires manual operation. In the future, an automatic collection system would be developed to achieve the integration of data collection, data analysis, and result display. In addition, human activity recognition for multiple targets will cope with enormous challenges. Existing works are based on the activity recognition of a single person. However, in real-life scenarios, human activity is an intricate combination of many activity types. Therefore, it is necessary for us to recognize the human activity involving more targets from

the intricate human activities. Meanwhile, it is vital to further explore the mechanisms that meet different situations and promote the practical application of human activity recognition in social life.

8. Conclusions and Future Work

8.1. Conclusions. In this paper, a framework for human activity recognition was proposed to improve the speed and accuracy of activity recognition. The framework contained three modules, which were developed to remove insensitive antennas, extract the range of human activities, reduce computational costs, and process redundant information. First, by analyzing the sensitivity between different antennas, an antenna selection approach was proposed to deal with insensitive antennas. After that, we enhanced the extracted sensitive antenna signals and discussed two different signal enhancement approaches, which can clearly show the active range and the inactive range. Finally, an activity segmentation algorithm was proposed to determine the beginning and end of the activity.

In our paper, three impact factors are discussed, namely, the diversity of human activities, the value of p in PSE, and the size of the sliding window. Although some progress has been made in human activity recognition, there still exist some challenging problems in our future work. We will continue to explore these problems and look forward to achieving satisfactory results.

8.2. Future Work

8.2.1. To Achieve Multiperson Recognition. In real-life scenarios, it is possible that multiple targets simultaneously exist in the same environment, and the activities are intertwined and complicated. Therefore, more equipment will be used to emulate the real-life situation in the following studies to obtain more signal information reflected by the human body. Meanwhile, the framework proposed in this paper will be applied to multiperson recognition and detection. Of course, the further research is challenging.

8.2.2. To Achieve Automatic Data Collection and Analysis. The existing data collection requires manual operation, which will cause additional interference by nonidentity personnel. In the future, we consider establishing a human activity recognition system which can perform the collection and analysis of signals automatically. The system is desired to achieve real-time activity recognition, visualization, and alarm.

8.2.3. To Introduce New Features. The existing features are based on empirical observation and statistical learning. It heavily depends on the specific environment which is deployed in our experiments. However, many factors, such as different environments, different individuals, and even different positions of the same individual, contribute to the accuracy. Our future researches will extend the study scope and depth for the features of WiFi signal and explore more effective methods of human activity recognition, which can be widely used in social life.

Data Availability

The CSI data for human behavior recognition used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-home gesture recognition using wireless signals," in *Proceedings of the ACM SIGCOMM 2013 Conference on SIGCOMM*, Hong Kong, China, August 2013.

[2] X. Guo, D. Zhang, K. Wu, and L. M. Ni, "MODLoc: localizing multiple objects in dynamic indoor environment," *IEEE Transactions on Parallel & Distributed Systems*, vol. 25, 2014.

[3] Z. Jin, W. Bo, H. Wen, S. S. Kanhere, and A. Tan, "Human identification using WiFi signal," in *Proceedings of the IEEE International Conference on Pervasive Computing & Communication Workshops*, Sydney, Australia, March 2016.

[4] A. Nandy, J. Saha, and C. Chowdhury, "Novel features for intensive human activity recognition based on wearable and smartphone sensors," *Microsystem Technologies*, vol. 26, no. 6, pp. 1889–1903, 2020.

[5] A. A. Núñez-Marcos, G. Azkune, and I. Arganda-Carreras, "Vision-based fall detection with convolutional neural networks," *Wireless Communications & Mobile Computing*, vol. 2017, Article ID 9474806, 16 pages, 2017.

[6] N. Okumura, Y. Yamanoi, R. Kato, and O. Yamamura, "Fall detection and walking estimation using floor vibration for solitary elderly people," in *Proceedings of the 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, Bari, Italy, October 2019.

[7] F. Adib and D. Katabi, "See through walls with WiFi!" *ACM SIGCOMM Computer Communication Review*, vol. 43, no. 4, pp. 75–86, 2013.

[8] H. Abdelnasser, K. A. Harras, and M. Youssef, "WiGest demo: a ubiquitous WiFi-based gesture recognition system," in *Proceedings of the 2015 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, Hong Kong, China, April 2015.

[9] M. T. Hoang, B. Yuen, X. Dong, T. Lu, R. Westendorp, and K. Reddy, "Recurrent neural networks for accurate RSSI indoor localization," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 10639–10651, 2019.

[10] Y. Gu, L. Quan, and F. Ren, "WiFi-assisted human activity recognition," in *Proceedings of the IEEE Asia Pacific Conference on Wireless & Mobile*, Bali, Indonesia, August 2014.

[11] J. Liu, L. Wang, J. Fang, L. Guo, B. Lu, and L. Shu, "Multi-target intense human motion analysis and detection using channel state information," *Sensors*, vol. 18, no. 10, 2018.

[12] J. Gjengset, J. Xiong, G. Mcphillips, and K. Jamieson, "Phaser: enabling phased array signal processing on commodity WiFi access points," in *Proceedings of the International Conference*

on Mobile Computing & Networking, Maui, HI, USA, May 2014.

[13] C. Xu, B. Firner, R. S. Moore, Y. Zhang, and N. An, "SCPL: indoor device-free multi-subject counting and localization using radio signal strength," in *Proceedings of the 12th International Conference on Information Processing in Sensor Networks—IPSN '13*, Philadelphia, PA, USA, April 2013.

[14] S. Liu, Y. Zhao, and B. Chen, "WiCount: a deep learning approach for crowd counting using WiFi signals," in *Proceedings of the 2017 IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC)*, Guangzhou, China, December 2017.

[15] K. Wu, J. Xiao, Y. Yi, M. Gao, and L. M. Ni, "FILA: fine-grained indoor localization," in *Proceedings of the IEEE INFOCOM*, pp. 2210–2218, Orlando, FL, USA, May 2012.

[16] C. Han, K. Wu, Y. Wang, and L. M. Ni, "WiFall: device-free fall detection by wireless networks," in *Proceedings of the IEEE INFOCOM 2014 IEEE Conference on Computer Communications*, Toronto, Canada, April 2014.

[17] W. Yan, L. Jian, Y. Chen, M. Gruteser, and H. Liu, "E-eyes: device-free location-oriented activity identification using fine-grained WiFi signatures," in *Proceedings of the 20th annual international conference on Mobile computing and networking*, Maui, HI, USA, September 2014.

[18] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Device-free human activity recognition using commercial WiFi devices," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 5, pp. 1118–1131, 2017.

[19] G. Linlin, W. Lei, L. Jialin, Z. Wei, and L. Bingxian, "HuAc: human activity recognition using crowdsourced WiFi signals and skeleton data," *Wireless Communications & Mobile Computing*, vol. 2018, Article ID 6163475, 15 pages, 2018.

[20] H. Yuan, X. Yang, A. He, Z. Li, and Z. Tian, "Features extraction and analysis for device-free human activity recognition based on channel statement information in b5G wireless communications," *EURASIP Journal on Wireless Communications and Networking*, vol. 2020, no. 1, 2020.

[21] X. Yang, R. Cao, M. Zhou, and L. Xie, "Temporal-frequency attention based human activity recognition using commercial WiFi devices," *IEEE Access*, vol. 8, pp. 137758–137769, 2020.

[22] J. Ding and Y. Wang, "WiFi CSI based human activity recognition using deep recurrent neural network," *IEEE Access*, vol. 7, pp. 174257–174269, 2019.

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

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

[25] Y. Taso, S. C. Yeh, Y. Y. Liang, C. H. Wang, and S. H. Fang, "Subcarrier selection for efficient CSI-based indoor localization," *IOP Conference Series: Materials Science and Engineering*, vol. 383, Article ID 012017, 2018.

[26] Y. Xu, W. Yang, J. Wang, X. Zhou, H. Li, and L. Huang, "WiStep," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 4, pp. 1–23, 2018.

[27] X. Liu, H. Chen, X. Jiang, J. Qian, and A. Pescape, "Wi-CR: human action counting and recognition with Wi-Fi signals," in *Proceedings of the 2019 4th International Conference on Computing, Communications and Security (ICCCS)*, Rome, Italy, October 2019.

[28] M. Raja, V. Ghaderi, and S. Sigg, "WiBot! In-vehicle behaviour and gesture recognition using wireless network edge," in *Proceedings of the 2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS)*, Vienna, Austria, July 2018.

[29] Z. Yang, Z. Zhou, and Y. Liu, "From RSSI to CSI," *ACM Computing Surveys*, vol. 46, no. 2, pp. 1–32, 2013.

[30] D. Halperirr, W. Hu, A. Sheth, and D. Wetherall, "Tool release: gathering 802.11n traces with channel state information," *ACM SIGCOMM Computer Communication Review*, vol. 41, no. 1, p. 53, 2011.

[31] 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, 2018.

[32] Z. Hao, Y. Duan, X. Dang, and T. Zhang, "CSI-HC: a WiFi-based indoor complex human motion recognition method," *Mobile Information Systems*, vol. 2020, Article ID 3185416, 20 pages, 2020.

[33] X. Dang, Y. Huang, Z. Hao, and X. Si, "PCA-Kalman: device-free indoor human behavior detection with commodity Wi-Fi," *EURASIP Journal on Wireless Communications and Networking*, vol. 2018, no. 1, 2018.

[34] S. Palipana, D. Rojas, P. Agrawal, and D. Pesch, "FallDeFi: ubiquitous fall detection using commodity Wi-Fi devices," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 4, pp. 1–25, 2017.

[35] 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 Transactions on Mobile Computing*, vol. 16, no. 2, pp. 511–526, 2017.

[36] 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*, Springer International Publishing, New York, NY, USA, 2015.

[37] P. L. Gatti, *Probability Theory and Mathematical Statistics for Engineers*, Pergamon Press, Oxford, UK, 1984.

[38] L. Shangguan, Z. Li, Z. Yang, M. Li, Y. Liu, and J. Han, "OTrack: towards order tracking for tags in mobile RFID systems," *IEEE Transactions on Parallel & Distributed Systems*, vol. 25, no. 8, pp. 2114–2125, 2014.