Wi-ID: WiFi-Based Identification System Using Rock-Paper-Scissors Hand Gestures

Zhiwen Zheng
Hefei University of Technology

Nan Yu
Nanjing University

Jingyang Zhang
University of Electronic Science and Technology of China

Haipeng Dai
Nanjing University

Qingshan Wang (✉ qswang@hfut.edu.cn)
Hefei University of Technology   https://orcid.org/0000-0003-4264-0180

Qi Wang
Hefei University of Technology

Research Article

Wi-ID: WiFi-Based Identification System Using Rock-Paper-Scissors Hand Gestures

**Zhiwen Zheng · Nan Yu ·
Jingyang Zhang · Haipeng Dai ·
Qingshan Wang · Qi Wang ·**

Abstract This paper proposes using a WiFi-based identification system, Wi-ID, to identify users from their unique hand gestures. Hand gestures from the popular game rock-paper-scissors are utilized for the system's user authentication commands. The whole feature of three hand gestures is extracted instead of the single gesture feature extracted by the existing methods. Dynamic time warping (DTW) is utilized to analyze the amplitude information in the time domain based on linear discriminant analysis (LDA), while extract amplitude kurtosis (AP-KU) and shape skewness (SP-SK) are utilized to analyze the Wi-Fi signals energy distribution in the frequency domain. Based on the contributions of the extracted features, the random forests algorithm is utilized for weight inputs in the LSTM model. The experiment is conducted on a computer installed with an Intel 5300 wireless networking card to evaluate

✉ Qingshan Wang
School of Mathematics, Hefei University of Technology, 193 Tunxi Road, Hefei, Anhui province, China
E-mail: qswang@hfut.edu.cn

Zhiwen Zheng, Qi Wang
School of Mathematics, Hefei University of Technology, 193 Tunxi Road, Hefei, Anhui province, China
E-mail: 2018111283@mail.hfut.edu.cn, wangq@hfut.edu.cn

Nan Yu, Haipeng Dai
Department of Computer Science and Technology, Nanjing University, 163 Xianlin Avenue, Nanjing, Jiangsu province, China
E-mail: yunan19920714@163.com, haipengdai@nju.edu.cn

Jingyang Zhang
Department of Aeronautics and Astronautic, University of Electronic Science and Technology of China, 2006 Xiyuan Avenue, Chendu, Sichuan province, China
E-mail: jingyang_zhang2@163.com

Fig. 1 Wi-ID system scenarios

the effectiveness and robustness of the Wi-ID system. The experiment results showed the accuracy of the proposed Wi-ID system has a personal differentiation accuracy rate over 92%, and with an average accuracy of 96%. Authorized persons who performed incomplete hand gestures are identified with an accuracy of 92% and hostile intruders can be identified with a probability of 90%. Such performance demonstrates that the Wi-ID system achieved the aim of user authentication.

Keywords rock-paper-scissors · authentication · WiFi-based · LSTM

1 Introduction

The development of the Internet of Things (IoT) in recent years has resulted in the rapidly growing usage of mobile devices that has led to rising demands for reliable and convenient user authentication processes more and more important. IoT devices include, but are not limited to, smartphones, smartwatches, and wearable devices. User authentication is becoming one of the most common application scenarios on the Internet of Things, attracting a large number of researchers [1–3, 16, 36–39]. WiFi signals have several advantages [4] over cameras[5] and wearable devices[6], such as it is unaffected by lighting conditions, provides better identification range, as WiFi signal can penetrate physical barriers [7–10]. It's possible to use wifi signals to identify and authorize employees's access to smart office buildings, as shown in Fig. 1. When a person makes a hand gesture in the appropriate position of the entrance, the WiFi

device can identify the user based on their actions. Existing smart office applications utilize cameras to verify users' identity, however, vision-based methods are often limited by lighting conditions. In addition, these systems can easily invade or reveal the personal information of users. In contrast, WiFi signals are hardly affected by the environment such as light and not invasive to the users' privacy, which makes it possible to be utilized for user authentication.

There are a few works on WiFi-based identification systems. Lu et al. [4] proposed a continuous user authentication system to capture the unique physiological characteristics of human respiratory movements by WiFi signal. However, the range of breathing movement is very limited, which is relatively easy to be disturbed. Thus, the action with a large range of motion can be designed to reduce the interference caused by other objects in the environment.

In this paper, we propose the creation of a Wi-ID system, a WiFi-based identification system using gesture actions and utilizes Commercial-Off-The-Shelf (COTS) WiFi devices to verify user identity. The advantages of the proposed Wi-ID system over existing user authentication methods primarily include easily performing specific gestures and a secure identification process. However, there are two major challenges: how to design simple gestures to identify individuals easily? And how to extract features with high accuracy to isolate individuals from tiny WiFi signal variations? To solve the first challenge, the three gestures from the traditional rock-paper-scissors game are utilized for user authentication. This simple children's game, which originated from Han Dynasty in China, has been widely known in the world. In addition, these gestures reflect dynamic changes of structure and space when the hand performs a series of movements. These changes affect the transmission of WiFi signals to varying degrees, and these changes can be used to realize user authentication.

We then extract the whole feature of three hand gestures from both the time and frequency domains instead of the single gesture feature extracted. And we utilize LDA to convert coordinates in the time domain and extract DTW features to analyze the amplitude information. In the frequency domain, features AP-KU and SP-SK are extracted to analyze the Wi-Fi signal energy distribution information. The identification model is based on the extracted DTW, AP-KU, SP-SK, and other features. Random forest (RF) is utilized for applying weight to the model, and identify users through the LSTM cyclic neural network to realize user authentication.

The main contributions of this paper are summarized in the following:

- Hand gestures from the traditional rock-paper-scissors game are selected to verify the users' identity. These gestures include an open hand, a clenched fist, and a partially opened fist with two fingers extended. These gestures reflect the structural and spatial changes of these hand movements. Three other gestures are selected for comparison effective authenticators.
- The whole of features of three hand gestures instead of the single gesture feature extracted by the existing methods. We select DTW time-domain features based on linear discriminant analysis to analyze the amplitude in-

formation of gestures. Energy distribution information is analyzed by using frequency domain characteristics of WiFi signals, including AP-KU and SP-SK. In addition, radio frequency is applied to the data to distinguish the contribution of extracted features to the whole user authentication system.

– The effectiveness and robustness of the Wi-ID system are evaluated on a computer installed with an Intel Link 5300 wireless networking card. The experiment results show the accuracy of individual Wi-ID system differentiation is over 92% with an average accuracy rate of 96%. The Wi-ID system could also authenticate users based on incomplete hand gestures and identify hostile intruders from imitating hand gestures.

The rest of this paper is organized as follows: Section 2 of this paper illustrates related research, the CSI signals are reviewed in detail in Section 3, and the system design is described in Section 4. Section 5 describes the results and comprehensive evaluation of the user authentication experiment, and the paper is summarized in Section 6.

2 Related Work

Gesture recognition is the foundation of user authentication, which can be categorized into sensor-based, vision-based, and radio frequency-based human motion recognition systems.

2.1 Sensor-Based Motion Recognition

Sensor-based gesture recognition methods utilize wearable sensors to acquire features of individual activities that can be utilized to identify human motion [1–3,6,14,15]. Duan et al. [14] proposed a new gesture recognition system under the premise of limiting the number of EMG signal sensors. The three signal channels in the system can classify nine different gestures. He et al. [6] compared the performance of a single-channel ultrasound system using surface EMG signals, indicating the possibility of combining two signal patterns to perform gesture recognition. Mantyjarvi et al. [15] studied the utilization of gait signals acquired by the three-dimensional accelerometer carried by the user to identify the action. Correlation, frequency domain, and data distribution statistics are all utilized in this system. Using this new gait recognition method proves that it is feasible to identify motion. However, these methods require the sensor to be worn in a specific manner to ensure accurate operation. It is also important to note that not everyone is willing to wear the sensor.

2.2 Vision-Based Motion Recognition

The vision-based motion recognition methods utilize images and videos to recognize human activities [40]. Zimmermann et al. [5] proposed a method for estimating the three-dimensional position of a hand using conventional RGB images. These methods utilize a deep network-implicit three-dimensional connection combined with the key points detected in the image to estimate a clear three-dimensional profile. Chen [17] proposed a fusion approach for improving human action recognition based on two differing modality sensors consisting of a depth camera and an inertial body sensor. In [18], the researchers utilized body language to study automatic emotion recognition and define a complete framework for automatic emotional body gesture recognition. Shotton et al. [19] proposed a new method for quickly and accurately predicting the 3D position of humans using a single depth image, and designed an intermediate model of the human body to transform the difficult statistical problem of analyzing human postures into a simplified pixel classification. Moddrop [20] is based on multi-scale and multi-modal deep learning for gesture detection and localization. Photo Sleuth [48] is a web-based platform for user authentication, which can help users successfully identify unknown portraits. Song[49] utilized a single stereo camera to track the body and hands and adopt model-based methods and particle filters to reconstruct body posture in 3D space to achieve human-computer interaction. Each visual form captures spatial information at a specific spatial scale (such as half-length or at the hand movement level), and the entire system operates on three-time scales. Although the above vision-based methods have strong identification accuracy to a certain extent, from a privacy perspective, these methods are invasive in both office and smart home applications.

2.3 Radio Frequency-Based Motion Recognition

RF-based sensor technology has received widespread attention from many researchers due to the increasing popularity of wireless devices [21–35]. In [41], a dedicated hardware sensor that utilizes human body features to sense body movements serves as an antenna to sense the posture of the entire body. WiTrack [42] utilizes FMCW (Frequency Modulated Continuous Wave) technology to calculate the Time of Flight (TOF) to track the target as well as the user's three-dimensional motion. Li et al. [33] proposed an efficient method based on the seam carving algorithm, which can extract features of different gestures and users effectively to realize gesture recognition. Pu et al. [43] proposed the WiSee system, based on the USRP-N210 experimental platform, which utilized the Doppler effect of radio waves to realize gesture recognition. Kellogg et al. [44] designed a special low-overhead system, called AllSee, extracting gesture information from the TV and RFID signal with minimal power drain. Widar [11] is a Wi-Fi-based zero-effort cross-domain gesture recognition system. Widar is a one-fits-all model that requires only one-time training

but can adapt to different data domains. Gu et al. [12] proposed a device-free real-time behavior analysis system based on WiFi signals, which utilized signal distortions caused by gestures to interpret different gesture combinations. Chen et al. [13] extracted human gestures from Channel State Information (CSI) signals and trained features using an attention-based bidirectional Long Short-Term Memory (LSTM) network, to recognize the movements. In the above wireless radio-based systems, the high costs of hardware and regulations limit their practicality. Therefore, researchers have been focused on designing human motion tracing systems based on CSI signals that have been collected by low-cost and popular WiFi chipset. WiGest [45] utilized WiFi signal strength changes to recognize a series of gestures. Wang et al.[46] analyzed the radio propagation model and proposed a fall detection system called WiFall. CARM [47] is a CSI-based human activity recognition system that establishes the correlation between CSI value dynamics, then CARM matches the given activities to the most suitable profile to recognize human activities. All the above studies focus on recognizing human motion and action based on CSI signals using commercial network cards, but none of these studies can identify people, since this requires a high degree of user authentication accuracy. Liu et al. [4] proposed a continuous user authentication system that extracts unique human respiratory biometrics from existing WiFi signals. This study takes advantage of WiFi-based signals the rock-paper-scissors hand gestures from traditional Chinese games to user authentication, which is robust to the interference caused by other objects for the gestures with a larger range of motion.

3 Channel State Information

CSI is a kind of data format for indicating the channel frequency response (CFR) of a sub-carrier granularity obtained by commercial IEEE 802.11a/g/n wireless network cards based on OFDM technology. Even with the same gestures, different users may exhibit subtle differences on their wireless channel due to their unique physiological characteristics (such as hand size and thickness) and behavioral characteristics (such as hand movements). Specifically, each group of CSI signals represents the amplitude and phase of an orthogonal frequency division multiplexing subcarrier, as follows

$$H(k) = ||H(k)||e^{j\angle H(k)}, \tag{1}$$

where $H(k)$ is the CSI measurement of subcarrier k, $||H(k)||$ and $\angle H(k)$ is the amplitude and phase of CSI measurement in subcarrier k, respectively.

Let N_t and N_r be the number of transmitting and receiving antennas. And, the MIMO system consists of $N_t \times N_r$ antenna pairs. The received signal can be expressed as

$$Y_k^p = H_k^p \times X_k^p + N k \in [1, C], p \in [1, N_t \times N_r], \tag{2}$$
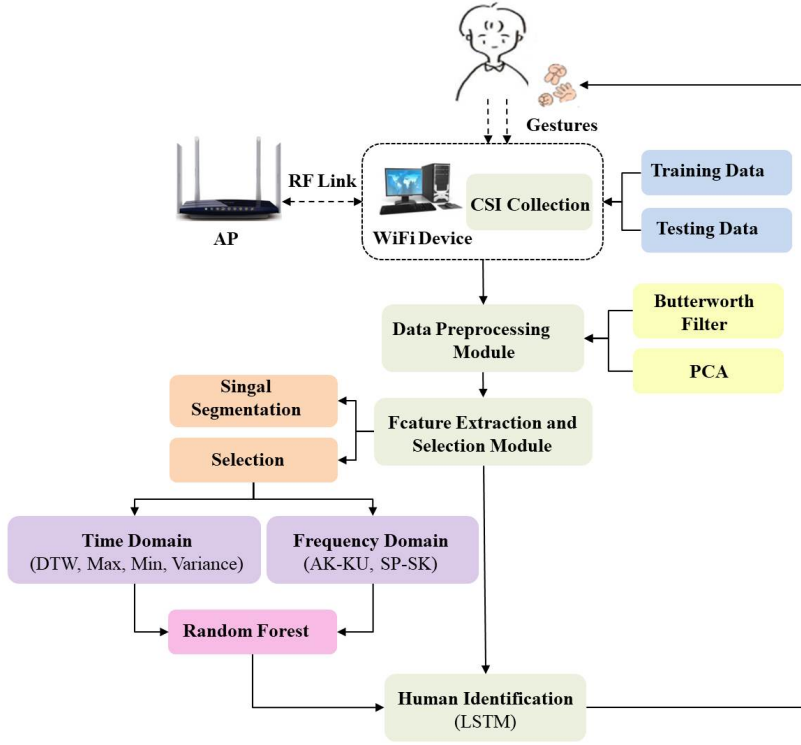
Fig. 2 Wi-ID system overview

where H_k^p represents channel frequency response (CFR) of antenna pair k on subcarrier p at any time. In this paper, we only focus on the amplitude of CSI measurement due to the imperfect phase of CSI measurement. The time stream of CSI measurement H_k^p is called the CSI stream. X_k^p denotes a signal vector of subcarrier p of antenna pair k. N is a noise vector, Gaussian White Noise. There are thirty 802.11n OFDM subcarrier settings in our experiment, therefore, we can obtain $C = 30$, and the total dimension of the CSI time series is $30 \times N_t \times N_r$.

4 System Design

The Wi-ID system consists of three modules: a data preprocessing module, feature extraction and selection module, and a human identity authentication module as shown in Fig. 2. Detailed descriptions about each module as follows:

4.1 System Overview

The proposed Wi-ID system is a <u>Wi</u>Fi-based <u>id</u>entification system using gesture actions, shown in Figure 2. The basic system concept is to identify individuals by using WiFi signals to capture individuals' unique signatures in a specified set of physical gestures. The data preprocessing module is illustrated in Fig. 2. Due to changes in the environment and ambient radio interference, the CSI measurement values collected from the environment are often noisy. To reduce environmental noise and interference, Principal Component Analysis (PCA) is utilized to process raw CSI measurements. In the feature extraction and selection module as shown in Fig. 2, we set the threshold of CSI measurement variance to determine the starting and ending point of the rock-paper-scissors gestures. In each feature extraction window, we extract gesture features in the time and frequency domains respectively. In the time domain, DTW is proposed to analyze amplitude information based on the LDA method. In the frequency domain, AP-KU and SP-SK are utilized to analyze energy distribution. In the human identity authentication module shown in Fig. 2, RF is utilized for applying weights to the feature extraction data. Different weights are assigned to corresponding features extracted from the feature extraction and selection module to train the LSTM model. The Wi-ID system identifies the user based on the training model.

4.2 Data Preprocessing

CSI measurement will be interfered with by non-testers behaviors and hardware errors. Thus, the Butterworth filter is utilized to denoise the original CSI stream. Given that the background noise frequency is much higher than any given gesture frequency, we utilize a Butterworth low-pass filter to remove the high-frequency environmental noise from the collected signals [14]. The frequency response of the Butterworth filter is the flattest filter in the passband and becomes zero in the stopband. Normalized Butterworth filters are defined in the frequency domain as follows:

$$|H_n(jw)| = \frac{1}{\sqrt{1 + w^{2n}}},$$ (3)

where n denotes the order of Butterworth filter, j represents an imaginary unit, and w is the angular frequency of signals in radians per second. w_c is utilized to denote the cut-off frequency of the Butterworth filter. The sampling rate F_s in the experiment is set at 1000 samples per second. Normally, the frequency of human gesture movement ranges from 0 to 15 Hz, with a median frequency of 5.125 GHz. Therefore, the cut-off frequency of the Butterworth low-pass filter is calculated as follows:

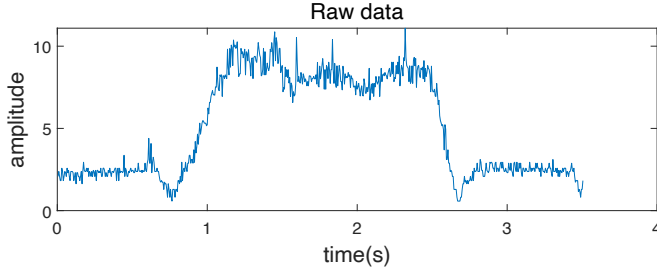$$w_c = \frac{2\pi f}{F_s} = \frac{2\pi 15}{1000} = 9.42.$$ (4)

Fig. 3 Raw CSI data

After applying the Butterworth filter, the high-frequency environmental noise is removed from the received signals. However, the simple low-pass Butterworth filter could not effectively filter interference from other actions, hence PCA is applied to further process the filtered data to obtain fine-granularity gesture information. CSI streams have multiple subcarriers, including 30 subcarriers in IEEE 802.11n, and there is a correlative relationship between the background noise among multiple subcarriers. From the research, it can be observed that human gestures are continuous, and different changes take place on different subcarriers with the increase of time. PCA is a dimension reduction technique, which transforms linearly related variables into almost linearly independent variables by orthogonal transformation and generates new features based on the original features to realize feature extraction. Based on these facts, PCA is utilized to filter Gauss white noise and interference from other actions to reduce the dimension of the CSI stream. Therefore, the available part of the signals is obtained and used as the main dimension and showed a relatively large energy level. And the white noise and interference signals are considered as other dimensions. Then, redundant signal dimensions are removed and only leave the signal's most usable portion. After performing data processing, background noise and interference from other actions are removed, and the fine-granularity CSI information is obtained to reflect changes in the users' rock-paper-scissors gestures.

Fig. 3 is the CSI signal of the 30th subcarriers before data preprocessing. Fig. 4 is the post-data preprocessing result of the CSI signal of the eighth volunteer's rock-paper-scissors gestures over for three seconds. Following data preprocessing shown in Fig. 4, gesture changes more clearly, and the fluctuation in noise and interference on the CSI stream is minimized.

Fig. 5 and Fig. 6 show variations in CSI amplitude extracted over time for the 30th subcarriers when two users make rock-paper-scissors gestures on an 802.11n WiFi device. The CSI amplitude showed different trends between two users, which proved that CSI can capture individuals' unique physiological and behavioral characteristics.

Fig. 7 and Fig. 8 show the results of PCA dimensionality reduction of data on the 30th subcarrier of user 7 and user 8 respectively, where the X-axis represents the time point of spectrum diagram calculation, the Y-axis
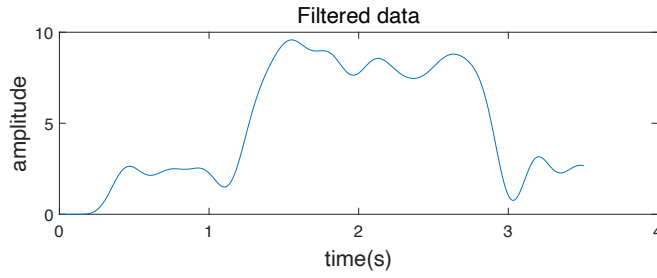
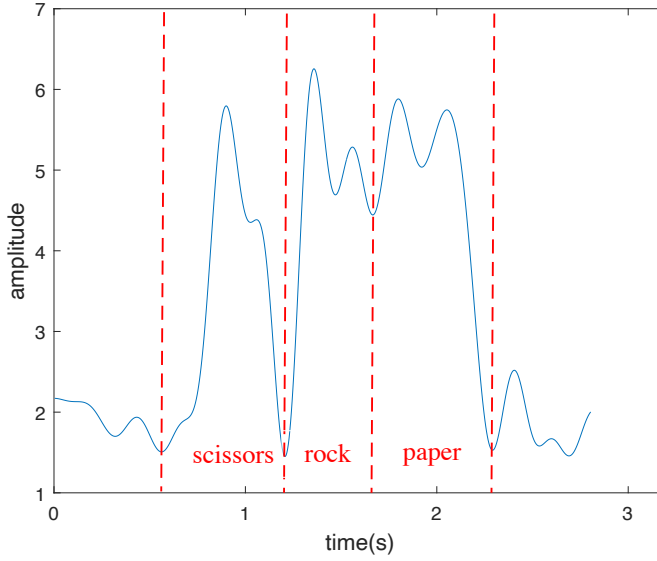**Fig. 4** CSI data after data preprocessing

Fig. 5 CSI amplitude of user 1

represents the signal frequency, and the Z-axis represents the PSD of this point. It can be observed the individuals' features are distributed across different locations after PCA dimension reduction.

4.3 Feature Extraction

In this section, we extract the features of the whole of gestures from the time domain and the frequency domains to realize user authentication. To provide finer granularity features for each activity, the CSI stream is first segmented to containing gesture movements into windows. Since there is a significant difference between the moving part and the stationary part of the CSI stream, the variance of CSI streams is utilized to segment continuous CSI signals. The characteristics of CSI signals in time domain and frequency domain are
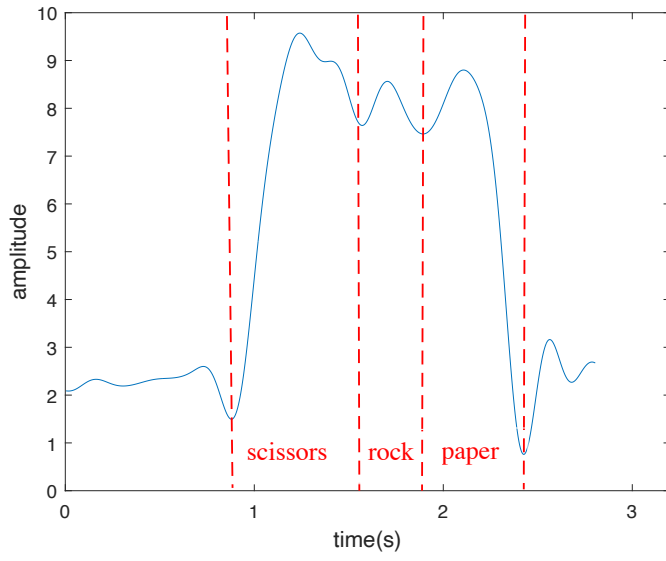
Fig. 6 CSI amplitude of user 2
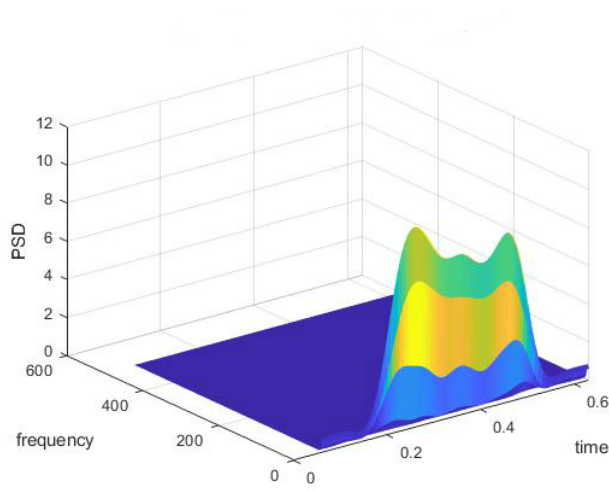
Fig. 7 PCA feature distribution for user 7

studied, and the amplitude information and energy distribution information of the CSI signals are analyzed respectively.

4.3.1 Feature Extraction at Time Domain

The Wi-ID system extracts the features of the whole three gestures in two steps in the time domain. In the first step, we utilized LDA to acquire new
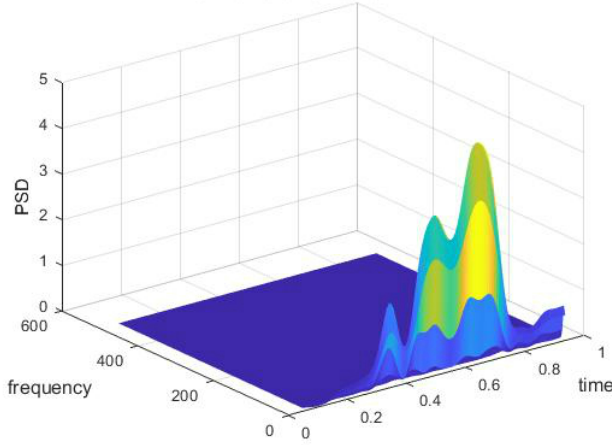
Fig. 8 PCA feature distribution for user 8

coordinates for the CSI stream. In the second step, DTW features are extracted based on the new coordinate system which is described in detail in the following.

Linear Discrimination Analysis (LDA) Technology: LDA is initially a dimensionality reduction technique for supervised learning to select the best projection for performance classification. In the proposed Wi-ID system, the CSI stream has a sequence of $\{(x_i, y_i)\}_{i=1}^{m}$, where m represents the CSI sequence length of CSI sequence. The LDA conversion process for the CSI sequence is described as follows. The CSI sequence is first projected into the line $y = \omega^T x$, where $x = [\{x_i\}_{i=1}^{m}]^T$ denotes a signal sequence's abscissa vector, and ω is a feature vector (column vector). Secondly, the covariance matrix of the same sample corresponding to the CSI signal sequence of a volunteer's gestures after projection is as follows:

$$\Sigma(\omega^T x - \omega^T u_i)^2 = \omega^T \Sigma[(x - u_i)(x - u_i)^T]w, \tag{5}$$

where u_i represents the mean of the CSI sequence, and $\Sigma[(x - u_i)(x - u_i)$ denotes the covariance matrix before the projection of CSI sequence of the same volunteer. Let Σ_0 and Σ_1 represent the covariance matrix before the projection of a pair of volunteer CSI sequences. The covariance matrices are obtained and projected to the above straight lines as $\omega^T \Sigma_0 \omega$ and $\omega^T \Sigma_1 \omega$, respectively. Finally, we minimize the covariance matrix of CSI sequences to make projection points fit more closely with the same volunteer. This is formulated as follows:

$$\min(\omega^T \Sigma_0 + \omega^T \Sigma_1 \omega). \tag{6}$$

Additionally, to differentiate volunteer projection points, the distance between different volunteer data centers must be as large as possible. This is formulated

as follows:

$$\max ||\omega^T u_0 - \omega^T u_1||. \tag{7}$$

By combining Equation (6) and Equation (7), the following optimization objectives are obtained:

$$\max J = \frac{||\omega^T u_0 - \omega^T u_1||}{\omega^T \Sigma_0 + \omega^T \Sigma_1 \omega} = \frac{\omega^T (u_0 - u_1)(u_0 - u_1)^T \omega}{\omega^T (\Sigma_0 + \Sigma_1)\omega}. \tag{8}$$

Dynamic Time Warping (DTW) Technology: The Wi-ID system obtained DTW features in the view of a new coordinate. The DTW feature method consists of two steps. Some notion for further analysis is first defined. Let the CSI sequences on channel k of two volunteers be A_k and B_k, respectively. Let the two volunteers CSI lengths be m and n, respectively. The specific expression is as follows:

$$A_k = a_{k1}, \cdots, a_{ki}, \cdots, a_{km}, \tag{9}$$

$$B_k = b_{k1}, \cdots, b_{kj}, \cdots, a_{kn}, \tag{10}$$

where $a_{ki} = (x_{ki}, y_{ki})$ represent the coordinates. The new coordinates for A_k and B_k using LDA can be represented as:

$$A_k^{'} = a_{k1}^{'}, \cdots, a_{ki}^{'}, \cdots, a_{km}^{'}, \tag{11}$$

$$B_k^{'} = b_{k1}^{'}, \cdots, b_{kj}^{'}, \cdots, a_{kn}^{'}. \tag{12}$$

Secondly, by letting DTW_k denote the DTW value of channel k between $a_{ki}^{'}$ and $b_{kj}^{'}$, the obtained results as follows:

$$DTW_k(i,j) = d(a_{ki}^{'}, b_{kj}^{'}) + \min\{DTW_k(i-1, j-1), DTW_k(i-1, j), DTW_k(i, j-1)\}, \tag{13}$$

where $d(a_{ki}^{'}, b_{kj}^{'})$ is the distance between $a_{ki}^{'}$ and $b_{kj}^{'}$ in the new coordinate system. In this paper, 30 channels are utilized. Therefore, the obtained DTW eigenvalue is $DTW = \frac{1}{30} \sum_{i=1}^{30} DTW_k$.

4.3.2 Feature Extraction at Frequency Domain

Considering the noise may disturb the features at the time domain, the features at the frequency domain for the whole of three gestures can be adopted for a better analysis. Therefore, the Wi-ID system also extracts features at the frequency domain. The Wi-ID system extracts two features at the frequency domain, that is, the PSD's AP-KU and SP-SK according to the different intensity of each user's gestures. AP-KU is calculated as follows:

$$AP - KU = \frac{1}{N} \sum_{i=1}^{N} [\frac{C(i) - \mu_{amp}}{\sigma_{amp}}]^4 - 3, \tag{14}$$

where

$$\mu_{amp} = \frac{1}{N}\sum_{i=1}^{N} C(i), \tag{15}$$

and

$$\sigma_{amp} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}[C(i) - \mu_{amp}]^2}, \tag{16}$$

where $C(i)$ is the frequency amplitude value of the window n, and N denotes the total number of windows. Considering the different positions and amplitude of each user's hand gesture motions, the SP-SK of the shape statistics is chosen as the primary feature. SP-SK can be formulated as follows:

$$SP - SK = \frac{1}{S}\sum_{i=1}^{N}[\frac{i - \mu_{amp}}{\sigma_{shape}}]^3 C(i), \tag{17}$$

where

$$\mu_{shape} = \frac{1}{S}\sum_{i=1}^{N} iC(i), \tag{18}$$

and

$$\sigma_{amp} = \sqrt{\frac{1}{S}\sum_{i=1}^{N}[i - \mu_{shape}]^2 C(i)}. \tag{19}$$

4.4 Classification and Identification

User authentication is fundamentally related to user security issues. Thus, Wi-ID system must have high recognition accuracy. We design an LSTM-based network to identify individuals. The DTW feature is extracted from the time domain, and features are also extracted from the maximum, minimum, and mean values of the second component after PCA. Thus, AP-KU and SP-SK values at the frequency domain are obtained. The Wi-ID system utilizes RF to assign different weights to different features to improve user authentication accuracy. Let $x_1, \cdots, x_6$ denote the six features extracted from the Wi-ID system. In each tree of the RF system, a tree is built to randomly extract autonomous training samples and calculated the out-of-bag (OOB) error rate. A very important feature would have a change that greatly impacts the test error, while an insignificant feature would have little impact on the test error. We utilize OOB samples to obtain error e_1, and randomly changed column j in OOB, but kept everything else the same. To realize the error, random vertical permutations are performed on column j to obtain error e_2. We characterize the importance of feature j as $e_1 - e_2$, normalized it, and averaged it to obtain the permutation importance index VIM_j^{OOB} of x_j. We can then obtain

$$VIM_{i,j}^{OOB} = \frac{\sum_{p=1}^{n_0^i} I(Y_p = Y_p^i)}{n_0^i} - \frac{\sum_{p=1}^{n_0^i} I(Y_p = Y_{p,\pi_i}^i)}{n_0^i}, \tag{20}$$
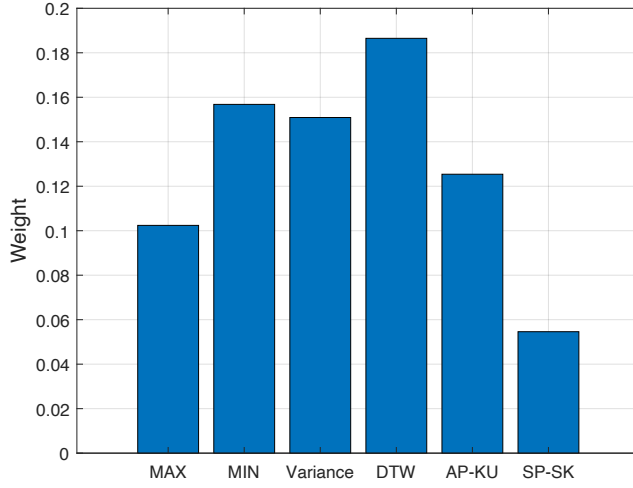
Fig. 9 Weight of all feature distribution

where n_0^i is the observation number outside the tree bag i, $I(.)$ is the indicator function, $Y_p \in \{0, 1\}$ is the observation result before the random permutations, changing column j data in OOB is denoted as π_j, and Y_{p,π_i}^i is prediction result of the i-th tree on the p-th observation of OOB data after random permutations. VIM_j^{OOB} of x_j ($1 \leq j \leq 6$) is obtained by taking the average value

$$VIM_j^{OOB} = \frac{\sum_{j=1}^{n} VIM_{i,j}^{OOB}}{n}. \tag{21}$$

Fig. 9 shows the feature weights of a data set from volunteers. The x-axis represents the selected feature class, and the y-axis represents the weight. It can be observed that the feature of the selected DTW has a higher weight than other features. In addition, the weight of the time domain features is often higher than frequency-domain features. The training design is based on the LSTM neural network. Specifically, the input vector of the LSTM is features extracted in the time domain and frequency domain of CSI data with 30 subcarriers and weighted by random forests, such as features obtained by LDA algorithm, AP-KU, SP-SK, and maximum and minimum of each CSI subcarrier.

5 Implementation and Evaluation

In this section, we conduct experiments to evaluate the performance of our proposed Wi-ID system.

5.1 Experimental Setup

Existing commercial hardware devices are utilized for implementing the Wi-ID system. Two PCs are utilized as transceivers, and the network card is an Intel 5300 wireless network card. Using Ubuntu version 14.04, the two computers are equipped with an Intel Core i9 processor and 8 GB RAM. They are equipped with one transmitting antenna and three receiving antennas. These antennas are placed horizontally above the ground. The distance between the transmitting antenna and the receiving antenna is 1.5 meters. The volunteer making gestures stands in the line between the transmitting antenna and the receiving antenna and is 0.2 meters from the transmitting antenna, which is shown as location 1 in Fig. 13. The transmission frequency of the transmission equipment is 1000 Hz. The experiment data sets[1] were collected from 30 volunteers (also referred to as users in this paper), including 15 males and 15 females. The volunteers are in the age range of 20 to 36 years old. To ensure the effectiveness and robustness of data acquisition, we collect the rock-paper-scissors gestures in the morning, noon, and night of a week. During one data collection session, the volunteers made a continuous rock-paper-scissors gesture for 5 seconds and repeated it for 10 minutes. We obtained a total of 37800 $(30 \times 7 \times 3 \times 60)$ gesture data samples from 30 volunteers, which were labeled according to user names. Among them, 80% of the samples were selected as the training set for model training, and the rest of them were selected as the validation set for model parameter adjustment.

5.2 Wi-ID system Accuracy

The system design aimed to identify volunteers from their rock-paper-scissors gestures. And to verify the validity of our model, we conducted the following experiments.

5.2.1 User Identification Accuracy

To verify our model on a test data set that is coming from the same setup of the training data set, we selected 20 volunteers randomly and asked them to do rock-paper-scissors gestures in 10 minutes as the test data set in turn under the same configuration environment. Fig. 10 shows the confusion matrix identified by these volunteers. The experimental results show that the identification accuracy is over 92% for these 20 volunteers, with an average accuracy of 96.86%. For user 6, user 11, and user 20, the identification accuracy is 100% because of their highly consistent gestures. Several volunteers' identification accuracy is reduced because of the errors they made while making gestures. Fig. 11 shows the accuracy rate, recall rate, and F1-score value of the Wi-ID

---

[1] Z. Zheng, J. Zhang, and Q. S. Wang, "Three gesture signals in WiFi environment," 2020, [Online] https://ieee-dataport.org/documents/three-gesture-signals-wifi-environment.

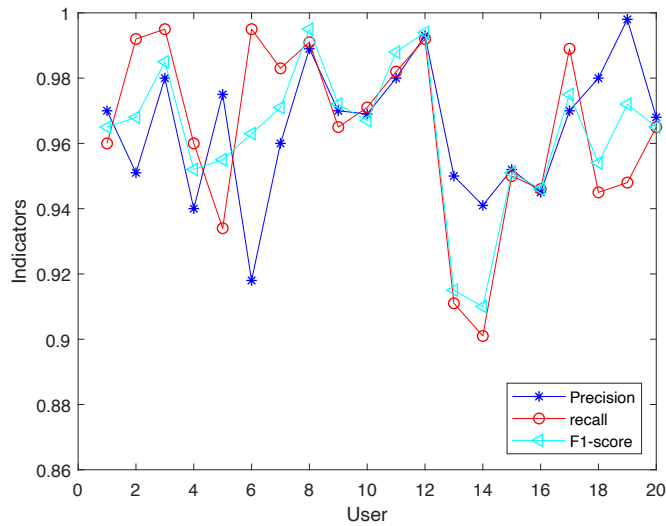**Fig. 10** Confusion matrix of user identification

Fig. 11 Precision/recall/F1-score with different users

system for selected 20 volunteers. Here, we can see that our proposed system can effectively identify volunteers in the general environment.
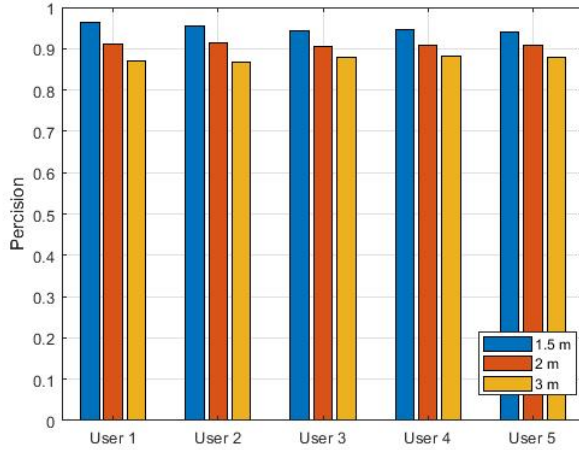
Fig. 12 Influence of distance between router and PC on identification effect

Table 1 Results of different distrubtion of people and transceivers

Experimental Setting	Average Precision
Position 1	94.98%
Position 2	92.53%

5.2.2 Accuracy in Different Experimental Setups

We conducted experiments on test data sets collected from different experimental setups with the training data sets to test the performance of the Wi-ID system.

Different Distances between Router and PC. To verify the impact of different distances between router and PC on the performance of the Wi-ID system, 5 volunteers are selected randomly to perform tests (three men and two women). In the experiment, the distance between the router and the PC was set at 1.5 m, 2 m, and 3 m, respectively. The identification effect at these distances is shown in Fig. 12, and these 5 volunteers all performed highly in all types of discriminant scenarios.

Different locations of volunteers To verify the impact of the different locations of volunteers on the Wi-ID system's performance, we changed the locations of volunteers on the experiment. We randomly selected 10 volunteers to perform the gestures on location 1 and location 2 in Fig. 13. As Fig. 13 shown, the distance between location 1 and location 2 is 1 m. Table 1 shows the Wi-ID system's average precision for location 1 and location 2 with these 10 volunteers, which proves that the Wi-ID system has a good performance for the test data sets collected from different locations.
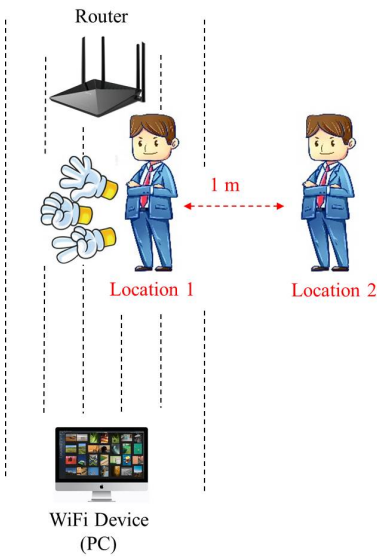
Fig. 13 The locations of volunteers

Fig. 14 Recognition result of rock-paper-scissors gesturs
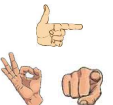
Fig. 15 Recognition result of "OK"-"number eight"-"pointing" gestures

Table 2 Incomplete gesture experimental results of user 2

User No.	Precision	User No.	Precision
User 1	0.00%	User 11	0.56%
User 2	92.08%	User 12	0.00%
User 3	0.08%	User 13	0.25%
User 4	2.25%	User 14	0.00%
User 5	1.04%	User 15	0.00%
User 6	0.00%	User 16	0.12%
User 7	1.46%	User 17	0.00%
User 8	0.00%	User 18	0.00%
User 9	0.00%	User 19	0.24%
User 10	1.92%	User 20	0.00%

5.2.3 Accuracy for Different Gesture Sets

To verify the gesture set impact on the performance of the Wi-ID, we experimented on two different gesture sets for the Wi-ID. One is the chosen gesture set of rock, paper, and scissors by this paper. The other gesture set consists of the hand gestures for "OK", "number eight", and "pointing" gestures that are randomly selected from the nine gestures that are stated as classifiable in [14] through the three channels of the surface EMG signal. Two volunteers are selected to perform these two sets of gestures for 10 minutes, and the data is processed and sent to the Wi-ID system. Fig. 14 and Fig. 15 show the experimental results. It can be observed that the individual discrimination of the rock-paper-scissors gestures is greater than the other gesture set. Therefore, the rock-paper-scissors gestures can uniquely identify individuals to realize the goal of accurate user authentication.

5.2.4 Accuracy of Incomplete Gestures

Considering that users might only make partial gestures due to habits or other reasons, we also carried out an incomplete gesture recognition experiment where the user deliberately completes only two-thirds of the entire rock-paper-and scissor gestures instead of the complete gesture. One out of the 20 volunteers is randomly chosen to perform incomplete rock-paper-scissors gestures for 3 minutes to obtain a test data set, that then ran in the proposed Wi-ID system. Table 2 shows the experimental results of the proposed Wi-ID system. The randomly chosen person in the experiment is user 2, and Table 2 clearly shows that our Wi-ID system could still identify user 2 based on the person's incomplete gestures. Thus our system can still verify user identity in case of incomplete gestures.

Table 3 Results of robustness for preventing intruders

User No.	Precision	User No.	Precision
User 1	0.00%	User 11	0.00%
User 2	90.05%	User 12	0.32%
User 3	0.00%	User 13	0.28%
User 4	0.45%	User 14	0.00%
User 5	7.95%	User 15	0.00%
User 6	0.00%	User 16	0.12%
User 7	0.08%	User 17	0.00%
User 8	0.00%	User 18	0.22%
User 9	0.00%	User 19	0.08%
User 10	0.45%	User 20	0.00%

Table 4 Experimental results of robustness for multiple interference sources

Experimental Setting	Average Precision
Without Multiple Interference Sources	94.98%
Multiple Interference Sources	83.21%

5.3 Robustness of Wi-ID

5.3.1 Robustness in preventing Intruders

The proposed WI-ID system must take into account the possibility of malicious intruders trying to force their way into environments, such as smart homes or smart offices by imitating hand gestures. Therefore, a simulated intrusion experiment is conducted to verify whether the system can accurately identify simulated intrusion attempts by non-authorized users. Out of the 20 volunteers, two volunteers with the most similar physical characteristics (such as hand size and thickness) are selected. User 2 and User 5 are selected. User 5 is chosen as the main subject while user 2 simulated an intrusion attempt. The collected signals are run through the Wi-ID system, and the identification results are obtained. Table 3 shows the results of the simulated intrusion experiment. From Table 3 it can be seen that the frequency where the Wi-ID system misidentified user 2 as user 5 is low, which proves that the Wi-ID system can still accurately identify individuals for user authentication purposes under hostile intrusion conditions.

5.3.2 Robustness against Multiple Interference Sources

To verify the Wi-ID system identification effect from other people's interference, we added a few bystanders to pace back and forth a meter away from a user undergoing the user authentication process. Table 4 shows the identification result of this experiment. From the results it can be observed even with the presence of human interference, the Wi-ID system is still reliable and still has high identification accuracy.

6 Conclusion

In this paper, we have proposed the creation of a Wi-ID system, a WiFi-based device-free personal identification system. Rock-paper-scissors gestures are selected as input commands for the Wi-ID system for this study, as the gestures involve individualized spatial changes between fingers, palms, and the backs of the hands. For data processing, Butterworth low-pass filtering and PCA are applied to reduce noise in the CSI data. For feature extraction, the whole feature of three gestures is extracted from the time domain and the frequency domain respectively. In particular, a new feature DTW is utilized in the time domain after coordinate transformation is performed with LDA, and AP-KU and SP-SK features are extracted from the frequency domain. Based on the contribution of the extracted features to the whole model, this paper utilizes random forests to assign different weights to the feature value, and the LSTM model is designed to identify individuals. Experimental results show that the identification accuracy of the Wi-ID system is more than 92% for different users. In terms of robustness, the Wi-ID system can accurately identify malicious imitation behaviors of illegal users and achieve high identification accuracy at different distances between the router and PC human interference conditions.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Yavuz, A. A., & Ozmen, M. O. (2019). Ultra lightweight multiple-time digital signature for the Internet of things devices. *IEEE Trans. Serv. Comput..* https://doi.org/10.1109/TSC.2019.2928303
2. Jain, A. K., & Arora, S. S. (2017). Fingerprint recognition of young children. *IEEE Trans. inf. Forensics Security,* 12(7), 1501-1514.
3. He, R., Wu, X., Sun, Z. N., & Tan, T. N. (2019). Wasserstein cnn: learning invariant features for nir-vis face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.,* 41(7), 1761-1773.
4. Liu, J., Dong, Y. D., & Chen, Y. Y. (2018). Leveraging breathing for continuous user authentication. In *24th Annual International Conference on Mobile Computing & Networking (MobiCom).* ACM.
5. Zimmermann, C., & Brox, T. (2017). Learning to estimate 3d hand pose from single rgb images. In *IEEE International Conference on Computer Vision (ICCV)* (pp. 4903-4911). IEEE.
6. He, J. Y., & Luo, H. (2019). Wrist and finger gesture recognition with single-element ultrasound signals: a comparison with single-channel surface electromyogram. *IEEE Trans. Biomed. Eng.,* 66(5), 1277-1284.
7. Zeng, Y. W., Wu, D, Xiong, J. Yi, E. & Gao, R. Y. (2019). Farsense: pushing the range limit of wifi-based respiration sensing with csi ratio of two antennas. In *ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* (pp. 1-26). ACM.
8. Sheng, B. Y, Fang, Y. R, & Xiao, F. (2020). An accurate device-free action recognition system using two-stream network. *IEEE Trans. Veh. Technol.,* 69(7), 7930-7939.

9. Xiao, L., Wan, X. Y, & Han, Z. (2018). PHY-layer authentication with multiple land-marks with reduced overhead. *IEEE Trans. Wireless Commun.,* 17(3), 1676-1687.
10. Hua, J., Sun, H., Shen, Z., Qian, Z., & Zhong, S. (2018). Accurate and efficient wireless device fingerprinting using channel state information. In *International Conference on Computer Communications (INFOCOM)* (pp. 1700-1708). IEEE.
11. Zheng, Y., Zhang, Y., & Qian, K. (2019). Zero-effort cross-domain gesture recognition with Wi-Fi. In *17th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys)* (pp. 313-325). ACM.
12. Gu, Y., & Zhang, X. (2018). Your WiFi knows how you behave: leveraging WiFi channel data for behavior analysis. In *IEEE Global Communications Conference (GLOBECOM)* (pp. 1-6). IEEE.
13. Chen, Z. H., Zhang, L., & Jiang, C. Y. (2019). WiFi CSI based passive human activity recognition using attention based BLSTM. *IEEE Trans. Mobile Comput.,* 18(11), 2714-2724.
14. Duan, F., Ren, X., & Yang, Y. K. (2021). A gesture recognition system based on time domain features and linear discriminant analysis. *IEEE Trans. Cognitive and Developmental Systems (TCDS).,* 13(1), 200-208. https://doi.org/10.1109/TCDS.2018.2884942
15. Mantyjarvi, J., Lindholm, M., & Vildjiounaite, E. (2005). Identifying users of portable devices from gait pattern with accelerometers. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (pp. 973-976). IEEE.
16. Teh, T. Y., Lee, Y. S., Cheah, Z. Y. (2017). IBI-Mobile Authentication: A Prototype to Facilitate Access Control Using Identity-Based Identification on Mobile Smart Devices. *Wireless Pers. Commun.,* 94, 127–144.
17. Chen, C., Jafari, R., & Kehtarnavaz, N. (2015). Improving human action recognition using fusion of depth camera and inertial sensors. *IEEE Trans. Human-Mach. Syst.,* 45(1), 51-61. https://doi.org/10.1109/THMS.2014.2362520
18. Noroozi, F., Kaminska, D., & Corneanu, C. (2018). Survey on emotional body gesture recognition. *IEEE Trans. Affect. Comput.,* 12(2), 505-523. https://doi.org/10.1109/TAFFC.2018.2874986
19. Shotton, J., Fitzgibbon, A., & Cook, M. (2011). Real-time human pose recognition in parts from single depth images. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1297-1304). IEEE.
20. Neverova, N., Wolf, C., Taylor, G., & Nebout, F. (2016). Moddrop: adaptive multi-modal gesture recognition. *IEEE Trans. Pattern Anal. Mach. Intell.,* 38(8), 1692-1706. https://doi.org/10.1109/TPAMI.2015.2461544
21. Liu, H., Wang, Y., Liu, J., Yang, J., Chen, Y., & Poor, H. V. (2017). Authenticating users through fine-grained channel information. *IEEE Trans. Mobile Comput.,* 17(2), 251-264.
22. Chen, Y., Wang, W., & Zhang, Q. (2014). Privacy-preserving location authentication in WiFi with fine-grained physical layer information. In *IEEE Global Communications Conference (GLOBECOM)* (pp. 4827-4832). IEEE.
23. Bao, Y., Dong, L., Zheng, Y., & Liu, Y. (2019). WiSafe: a real-time system for intrusion detection based on wifi signals. In *the ACM Turing Celebration Conference-China (TURC)* (pp. 1-5). ACM.
24. Wang, F. X., & Gong, W. (2018). On spatial diversity in WiFi-based human activity recognition: A deep learning-based approach. *IEEE Internet of Things Journal.,* 6(2), 2035-2047.
25. Kong, H., Lu, L., Yu, J., Chen, Y., & Tang, F. (2020). Continuous authentication through finger gesture interaction for smart homes using WiFi. *IEEE Trans. Mobile Comput..* Advance online publication. https://doi.org/10.1109/TMC.2020.2994955
26. Yan, H., & Zhang, Y. (2020). WiAct: a passive WiFi-based human activity recognition system. *IEEE Sensors Journal.,* 20(1), 296-305. https://doi.org/10.1109/JSEN.2019.2938245
27. AliKhan, D., & Razak, S. (2019). Human behaviour recognition using Wifi channel state information. In *2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 7625-7629). IEEE.
28. Fei, H., Xiao, F., Han, J. S., Huang, H. P., & Sun, L. J. (2020). Multi-variations activity based gaits recognition using commodity WiFi. *IEEE Trans. Veh. Technol.,* 69(2), 2263-2273.

29. Wang, W., Alex, X., & Shahzad, M. (2016). Gait recognition using wifi signals. In *2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (IJCPUC)* (pp. 363-373). ACM.

30. Yu, N., Wang, W., Alex, X., & Kong, L. T. (2018). Qgesture: quantifying gesture distance and direction with wifi signals. In *the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* (pp. 1-23). ACM.

31. Qian, K., Wu, C. S., & Yang, Z. (2018). Enabling contactless detection of moving humans with dynamic speeds using CSI. *ACM Trans. Embed. Comput. Syst.,* 17(2), 1-18.

32. Abdelnasser, H., Harras, K., & Youssef, M. (2019). A ubiquitous WiFi-based fine-grained gesture recognition system. *IEEE Trans. Mobile Comput.,* 18(11), 2474-2487.

33. Li, C., Liu, M., & Cao, Z. (2020). WiHF: enable User Identified Gesture Recognition with WiFi. In *IEEE Conf. Comput. Commun. (INFOCOM)*. IEEE.

34. Zeng, Y., Wu, D., & Xiong, J. (2020). Boosting WiFi Sensing Performance via CSI Ratio. *IEEE Pervasive Comput.,* 20(1), 62-70. https://doi.org/10.1109/MPRV.2020.3041024

35. Wu, D., Gao, R., & Zeng, Y. (2020). Fingerdraw: Sub-wavelength level finger motion tracking with WiFi signals. In *ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* (pp. 1-27). ACM.

36. Anandhi, S., Anitha, R., & Sureshkumar, V. (2019). IoT Enabled RFID Authentication and Secure Object Tracking System for Smart Logistics. *Wireless Pers. Commun.,* 104, 543–560. https://doi.org/10.1007/s11277-018-6033-6

37. Lo, N. W., Yohan, A. (2020). BLE-Based Authentication Protocol for Micropayment Using Wearable Device. *Wireless Pers. Commun.,* 112, 2351–2372. https://doi.org/10.1007/s11277-020-07153-0

38. Ahn, J., Lee, I. G., & Kim, M. (2020). Design and Implementation of Hardware-Based Remote Attestation for a Secure Internet of Things. *Wireless Pers. Commun.,* 114, 295–327. https://doi.org/10.1007/s11277-020-07364-5

39. Li, X., Wen, Q., & Li, W. (2016). A Three-Factor Based Remote User Authentication Scheme: Strengthening Systematic Security and Personal Privacy for Wireless Communications. *Wireless Pers. Commun.,* 86, 1593–1610. https://doi.org/10.1007/s11277-015-3008-8

40. Muhammad, I. K., Mian, A. J., Yar, M., Dinh, T. D., Ateeq, R., Constandinos, X. M., & Evangelos, P. (2021). Tracking vital signs of a patient using channel state information and machine learning for a smart healthcare system. *Neural Comput. & Applic.*. Advance online publication. https://doi.org/10.1007/s00521-020-05631-x

41. Cohn, G., Morris, D., & Patel, S. (2012). Humantenna: using the body as an antenna for real-time whole-body interaction. In *the SIGCHI Conference on Human Factors in Computing Systems (CHI)* (pp. 1901-1910). ACM.

42. Zimmermann, C., & Brox, T. (2014). 3D tracking via body radio reflections. In *14th Networked Systems Design and Implementation (NSDI)* (pp. 317-329). ACM.

43. Pu, Q. F., Gupta, S., Gollakota, S., & Patel, S. (2013). Whole-home gesture recognition using wireless signals. In *19th Annual International Conference on Mobile Computing & Networking (MobiCom)* (pp. 27-38). ACM.

44. Kellogg, B., Talla, V., & Gollakota, S. (2014). Bringing gesture recognition to all devices. In *14th Networked Systems Design and Implementation (NSDI)* (pp. 303-316). ACM.

45. Abdelnasser, H., Youssef, M., & Harras, K. A. (2015). WiGest: a ubiquitous WiFi-based gesture recognition system. In *International Conference on Computer Communications (INFOCOM)* (pp. 1472-1480). IEEE.

46. Wang, Y. X., Wu, K. S., & Ni, L. M. (2016). Wifall: device-free fall detection by wireless networks. *IEEE Trans. Mobile Comput.,* 16(2), 581-594.

47. Wang, W., Liu, A. X., Shahzad, M., & Ling, K. (2015). Understanding and modeling of wifi signal based human activity recognition. In *21st Annual International Conference on Mobile Computing & Networking (MobiCom)* (pp. 65-76). ACM.

48. Mohanty, V., Thames, D., & Mehta, S. (2020). Photo Sleuth: Identifying Historical Portraits with Face Recognition and Crowdsourced Human Expertise. *ACM Trans. Interactive Intelligent Syst.,* 10(4), 1–36. https://doi.org/10.1145/3365842

49. Song, Y., Demirdjian, D., & Davis, R. (2012). Continuous body and hand gesture recognition for natural human-computer interaction. *ACM Trans. Interactive Intelligent Syst.,* 2(1), 1-28. https://doi.org/10.1145/2133366.2133371