ISSN: 2093-5374 / E-ISSN: 2093-5382

# WiFi-based Intelligent Wireless Sensing for Privacy-Preserving Human Behavior Recognition under AIoT Architecture

Haoda Wang[1], Liang Lin[2*], Huakun Huang[3], Lingjun Zhao[4], Zhuotao Lian[5], and Chunhua Su[6*]

[1]PhD Candidate, Graduate School of Computer Science and Engineering, The University of Aizu, Aizuwakamatsu, Japan. d8242105@u-aizu.ac.jp, https://orcid.org/0009-0009-8526-3988

[2*]Assistant Professor, Department of Information Engineering, Luoding Polytechnic, Yunfu, China. 13016054668@163.com, https://orcid.org/0000-0002-1161-5735

[3]Associate Professor, School of Computer Science and Cyber Engineering, Guangzhou University, Guangdong, China. huanghuakun@gzhu.edu.cn, https://orcid.org/0000-0003-2853-8892

[4]Associate Professor, Department of Network Engineering, School of Electronics and Information, Guangdong Polytechnic Normal University, Guangdong, China. atjonlyn@gmail.com, https://orcid.org/0000-0003-2369-8862

[5]Postdoctoral Fellow, Graduate School of Information Science and Electrical Engineering, Kyushu University, Fukuoka, Japan. zhuotaolian@ieee.org, https://orcid.org/0000-0003-2938-6368

[6*]Senior Associate Professor, Department of Computer Science and Engineering, Division of Computer Science, The University of Aizu, Aizuwakamatsu, Japan. chsu@u-aizu.ac.jp, https://orcid.org/0000-0002-6461-9684

Received: October 23, 2024; Revised: December 09, 2024; Accepted: January 10, 2025; Published: March 31, 2025

## Abstract

In the era of the Internet of Things (IoT) and AI-driven smart environments, human behavior recognition has emerged as a pivotal technology underpinning a broad spectrum of intelligent applications. However, achieving high recognition accuracy while preserving user privacy remains a critical challenge. To tackle this problem, this paper introduces a novel privacy-preserving method for WiFi-based human behavior recognition. It employs three-dimensional convolutional neural networks(3D-CNNs) enhanced with an attention-enabled autoencoder mechanism, called ThAN. The proposed approach also constructs three distinct classifiers activity, identity, and location that utilize 3D-CNNs for effective feature extraction from Channel State Information signals. Through rigorous experimentation with a real-world dataset, this study demonstrates that the method not only significantly safeguards privacy, achieving a high protection level, but also maintains a high recognition accuracy of 99%, substantiating its efficacy and applicability in real-world scenarios.

**Keywords:** Privacy Protection, Wireless Sensing, Human Behavior Recognition, Machine Learning, IoT.

*Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA)*, volume: 16, number: 1 (March), pp. 104-120. DOI: *10.58346/JOWUA.2025.I1.006*

[*]Corresponding author: Assistant Professor, Department of Information Engineering, Luoding Polytechnic, Yunfu, China.

Senior Associate Professor, Department of Computer Science and Engineering, Division of Computer Science, The University of Aizu, Aizuwakamatsu, Japan.

# 1 Introduction

The Internet of Things (IoT), as a technology that allows data exchange between IoT devices, has gained tremendous attention among researchers (Wang et al., 2024; Awaisi et al., 2024). It has a variety of implementations in various fields, such as smart homes, smart cities, visual reality (VR) interaction, entertainment, and healthcare. To enhance the efficiency of IoT, the integrated technology of Artificial Intelligence and IoT—termed AIoT—has emerged, leveraging the power of AI (Nwosu & Adeloye, 2023). Among these various applications, behavior recognition based on AIoT plays a pivotal role as it can smooth communication in VR social and healthcare domains and so on (Figure 1).

Hence, AIoT-enhanced human behavior recognition technology has attracted significant attention (Islam et al., 2023; Li et al., 2023), such as metaverse, VR social. Conventional research on human behavior recognition is focused on device-based technologies, which generally require carrying electronic devices or equipment, making it challenging to implement in some scenarios (Pan et al., 2024; Awadzi et al., 2018). For instance, elderly fall detection systems based on wearable devices fail if the elderly forget to wear or just refuse to wear (Hussain et al., 2020). To alleviate these constraints, researchers began to explore Device-free based human behavior recognition systems (Gupta et al., 2022; Yadav et al., 2022).

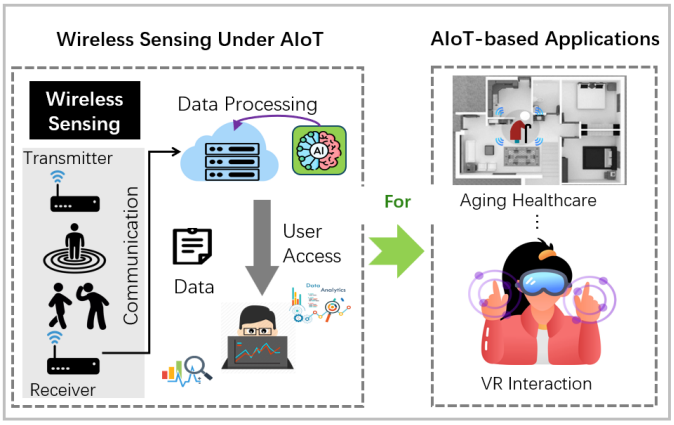



Figure 1: Architecture and Applications of Intelligent Wireless Sensing under the AIoT Environment

Device-free human behavior recognition systems can be categorized into several types based on their data collection IoT devices (Kavitha, 2024; Fridani & Vosoughi, 2014). Infrared sensing-based approaches commonly install infrared cameras in fixed locations and collect infrared radiation signals. Infrared sensing-based methods offer relatively high recognition accuracy, but they are easily affected by obstructions and temperature changes, and also present privacy concerns (Rezaei et al., 2022). Computer Vision-based methods use devices that include cameras and video recorders (Kavitha, 2024). By installing these devices in the target area, image and video data can be collected. Computer vision-based methods usually offer high recognition accuracy, but they are susceptible to some conditions such as lighting and obstructions (Klavin, 2024; Jeyanthi & Krishnamoorthi, 2022). Additionally, they involve privacy concerns and cannot be installed in spaces like restrooms or bathrooms. Other IoT device like radar-based methods implement devices such as UWB radar and millimeter-wave radar (Ahmed & Cho, 2023; Luo, 2023). It is possible to collect information including the amplitude, phase, and Doppler shift of reflected signals. This kind of method offers high recognition accuracy but has high costs, making it difficult to generalize.

Among these device-free methods, Wifi-based Wireless Sensing for human behavior recognition systems under AIoT architecture, however, offers the advantages of being more budget-friendly and less restrictive to limitations mentioned above (Abolqasem et al., 2015; Miki et al., 2024). The WiFi sensing-based method is not affected by lighting, allowing it to be used in dark environments. It can also penetrate walls, unaffected by obstructions, and can effectively recognize actions behind obstacles (Zhao et al., 2024; Miao et al., 2024). The cost of using methods based on wifi also remains low. Under wifi-based systems, channel state information (CSI) is widely used (Liang et al., 2023) because human behaviors result in unique variations at the CSI receivers.

Although AIoT enhanced Wireless sensing for human behavior recognition using CSI information demonstrates superior advantages such as low cost and high efficiency, it is still vulnerable to privacy leakage (Kavitha, 2024; Wang et al., 2022), especially in VR communication, due to the nature of data exchange in IoT devices (Clavijo-López et al., 2024). In WiFi sensing, the collected CSI information often includes sensitive or private content about individuals or targets, such as identity information, location data, and activity information. If hackers steal CSI data, they could analyze it to reveal identity, activity, and other private information. With growing concerns about privacy security, the leakage of customer privacy can also cause severe reputational and financial damage to service providers (Kumar, 2024). Hence, ensuring the privacy protection of user information has become a critical research topic. Nonetheless, researchers have primarily focused on improving recognition accuracy, with insufficient efforts made toward protecting privacy in WiFi sense.

Currently, the primary methods for privacy protection in AIoT Wi-Fisensing focus on the data exchange process, which involves modifying the transmitter's Channel State Information (CSI) signal and altering the CSI signal collected by the receiver. The latter mainly involves inputting data into autoencoder neural networks or adversarial neural networks (Chithra Devi et al., 2024). By setting specific loss functions, adversarial signals are generated to achieve privacy protection. Existing methods primarily use two-dimensional convolutional neural networks (2D-CNNs) to extract specific features before generating the required samples (Ghazi et al., 2021) [28]. However, CSI signals are video stream data, and the features extracted using 2D-CNNs are insufficient. The generated samples still contain some private information and reduce the accuracy of recognizing other information.

To solve this problem, this paper proposes, ThAN, a privacy protection method using three-dimensional convolutional neural networks (3D-CNNs) and a self-attention mechanism enabled autoencoder neural network. Experiments were conducted on real dataset (Shinan et al., 2024). The results show that the proposed algorithm achieves excellent privacy protection effects while significantly reducing its impact on other information recognition.

The main contributions of this paper can be summarized as follows:

- Three recognition classifiers were designed: an activity classifier, an identity classifier, and a location classifier. These classifiers utilize three-dimensional convolutional neural networks (3D-CNNs) to effectively extract temporal and spatial features from the data, thereby improving the recognition accuracy.
- An autoencoder neural network was designed using 3D-CNNs, transposed 3D-CNNs, and an attention mechanism. This network extracts features from sample data and generates adversarial samples, effectively protecting private information.
- The privacy protection ability and recognition performance of the proposed approach is verified on real-world datasets.

The remaining parts of this paper are organized as follows. Section 2 introduces the related works. Section 3 introduces the proposed algorithm design. Section 4 presents the experimental setting and performance evaluation. Section 5 concludes this work.

# 2 Related Work

### 1) Device-Free Human Behavior Recognition

With the rapid development of computer technology, recent research has demonstrated the potential of CSI in coarse-grained action recognition and fine-grained gesture recognition (Huang et al., 2023; Ding et al., 2023; Zhao et al., 2023). Currently, methods for action recognition using WiFi sensing are mainly divided into two types: action recognition using machine learning techniques and action recognition using deep learning techniques.

For WiFi-based human behavior recognition methods using machine learning techniques, Zhang et al., (2019) derived and extracted domain-independent features of human gestures at lower signal levels and developed a model that can adapt to different data domains with only one training session. This effectively enables cross-domain recognition without the need for retraining. Dang et al., (2020) integrated the dynamic time warping algorithm with the support vector machine (SVM) algorithm, achieving a significant improvement in action recognition accuracy through the joint extraction of amplitude and phase information. Xian et al., (2021) proposed a method combining the K-Means algorithm with dynamic time regularization, transforming feature-matching in gesture recognition into a classification problem, thereby effectively recognizing nine types of gestures. Moghaddam et al., (2023) proposed a method that combines CSI and RSS to enhance the performance of device-free fine-grained human activity recognition (HAR) using WiFi signals. They validated the approach using models such as support vector machines (SVM) and Gaussian Naive Bayes, achieving high accuracy in action recognition.

For WiFi-based human behavior recognition methods using deep learning techniques, Muaaz et al., (2022) employed the CSI ratio method to reduce noise and phase offset effects. Spectrograms were then computed for each action and used to train a CNN model, achieving a recognition accuracy of 97.78%. Duan et al., (2022) proposed a subcarrier correlation and inversion-based ranking algorithm to extract each user's signal. They combined a GRU with a one-dimensional convolutional neural network to recognize corresponding user actions. Meng et al., (2021) applied a sparse recovery method to reduce the dimensionality of CSI phase measurements and constructed a phase difference matrix (Atawneh et al., 2024). They proposed a bidirectional gated recurrent unit network based on an attention mechanism to automatically learn and extract discriminative features from the phase difference matrix. Meng et al., (2023) proposed a lightweight CSI-based human activity recognition (HAR) model that explores the graphical correlation of CSI subcarriers and combines it with a temporal causal convolution module, achieving efficient and accurate action recognition. Zhang et al., (2022) integrated CNN with a convolutional attention module, followed by bidirectional gated recurrent units and a self-attention mechanism to extract temporal features. Finally, an attention-based feature fusion module was employed to integrate semantic action features with temporal features, further improving recognition accuracy.

In summary, action recognition using machine learning methods faces several challenges. These methods require extensive prior knowledge, struggle to effectively distinguish similar actions, and demand high-quality data, limiting their ability to achieve high recognition accuracy. Deep learning methods can effectively extract spatial features; however, they fall short in adequately capturing temporal features, making it difficult to accurately recognize actions with reversed sequences. Therefore,

WiFi-based action recognition faces the challenge of better extracting temporal information from CSI signals.

### 2) Privacy Protection on Wifi Based Device Free HBR

As mentioned above, current primary privacy protection methods are modifying the transmitter's CSI signal or the receiver's collected CSI signal. Privacy protection methods on the transmitter side primarily involve modifying or perturbing the signals sent by the transmitter to prevent sensitive information from being leaked or exploited by malicious attackers during transmission. Abanto-Leon et al., (2020) introduced random phase errors into the transmitted signal, ensuring that only specific receiver devices could extract the transmitter's original signal. This method not only resists spoofing attacks but also confuses the transmitter's radiation signal, effectively protecting privacy. Jiao et al., (2021) embedded a CSI obfuscator in the WiFi transmitter, applying artificial channel responses to the signal before transmission to prevent unauthorized WiFi sensing without compromising WiFi link performance. Methods based on the transmitter require modifying the data before transmission. Previous studies achieved the effect of allowing only specific receivers to receive the signal, effectively preventing privacy leakage. However, these methods often require hardware modifications or may impact communication quality. Most research focuses primarily on protecting user identity information, while protection of action and location information has not been addressed.

Privacy protection methods based on the receiver primarily focus on modifying or processing the received signal to prevent sensitive information from being exposed during or after the reception process. Zhao et al., (2020) proposed a trusted fake location protection algorithm, where users locally generate trusted fake locations and use fake location queries derived from these trusted fake locations to hide their WiFi location queries. Zhang et al., (2022) proposed a privacy-preserving system framework based on a multi-level edge network. The framework employs differential privacy techniques, including private sample labeling obfuscation, differentially private feature fusion, and differentially private model training, for indoor edge computing-based localization. Zhang et al., (2020) proposed two types of adversarial autoencoder networks, which modify the protected semantic features by replacing the original signal with generated adversarial signals while preserving important features needed for other semantic recognition. Zhou et al., (2019) proposed a novel loss function for adversarial neural networks, enabling targeted modification of CSI signals to constrain new classification results to meet adversarial requirements. Receiver-based methods do not affect communication and are relatively easier to implement. Existing methods provide good privacy protection; however, they still have a significant impact on the recognition performance while protecting private information, which remains a major challenge.

## 3 The Proposed Algorithm ThAN

### 1) Preliminary

- **Convolutional Neural Network**

Convolutional Neural Networks (CNNs) are a type of deep neural network with a convolutional structure. Based on the dimensionality of the data they process, CNNs can be categorized into one-dimensional (1D-CNN), two-dimensional (2D-CNN), and three-dimensional (3D-CNN) convolutional neural networks.1D-CNN is primarily used for processing sequential data such as text, audio, and time series. It effectively captures local patterns and dependencies in the data, playing a significant role in fields

such as natural language processing, speech recognition, and audio analysis. 2D-CNNis mainly utilized in computer vision and image processing. It excels at capturing spatial features in images, such as edges, textures, and shapes, and performs remarkably well in tasks like image classification, object detection, and image generation. 3D-CNN first proposed (Ji et al., 2012) is designed to handle data with both spatial and temporal dimensions, such as video data. It has broad applications in areas like computer vision, medical image processing, and action recognition.

- **Self-Attention Mechanism**

The attention mechanism (Attention) is a method inspired by the human visual and cognitive system. By introducing the attention mechanism, neural networks can automatically focus on important information in the input, thereby improving model performance and training efficiency. The computation process of the attention mechanism can be divided into two stages: the first stage calculates the weight coefficients based on the query and key, and the second stage performs a weighted sum of the values using the weight coefficients. First, by linearly mapping the input matrix to different vector spaces, the query vector Q, key vector K, and value vector V can be obtained. This process can be described by Eq. (1), Eq. (2), and Eq. (3), where $W_q$,$W_k$, $W_v$ correspond to the linear mapping matrices for the query, key, and value, respectively.

| | |
|---|---|
| $Q = W_q X,$ | (1) |
| $K = W_q X,$ | (2) |
| $V = W_v X,$ | (3) |
| $Attention(Q,K,V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V,$ | (4) |

Next, the obtained $Q$ and $K$ are multiplied using the dot product. To prevent the kernel values from becoming too large, the result is divided by the square root of the input vector's dimension. Then, the Softmax function is applied to compute a weight matrix, where each element represents the correlation or similarity between the query and the key. Finally, by multiplying this weight matrix with V, important contextual information can be aggregated into the context vector based on these weight coefficients. The attention mechanism can be expressed using Eq. (4).

Self-attention is a special form of the attention mechanism, where the query, key, and value all come from the same input data. Self-attention allows the model to calculate scores within the data itself, improving the model's performance and expressive capability, especially for tasks involving long sequences or requiring finer control over specific parts of the input. It has been widely applied in fields such as natural language processing, machine translation, speech recognition, and image caption generation.

- **Autoencoder Neural Networks**

Autoencoder neural networks are a type of unsupervised learning algorithm primarily used for learning data representations. They can be applied to tasks such as dimensionality reduction, feature learning, and data generation. The structure of an autoencoder consists of two parts: the encoder and the decoder. Through the encoder, key features from the input data can be learned and compressed into a hidden space representation. Typically, the encoder is composed of fully connected layers or convolutional layers. By stacking multiple layers, the encoder extracts the main features from the data, reduces the dimensionality of the data or feature layers, and improves the accuracy of data generation. The extracted features are then passed through the decoder, which maps the learned features from the hidden space to a new space with the same dimensions as the input data. Generally, the decoder consists of fully

connected layers or transposed convolutional layers. By stacking multiple layers, the decoder gradually restores the dimensionality of the feature layers, ensuring that the generated data has the same dimensions as the input data. In this process, a loss function is used to adjust the difference between the input data and the generated data. Depending on the specific task, different structures of autoencoder neural networks can be designed, such as variational autoencoders (VAEs), denoising autoencoders, and spare autoencoders.

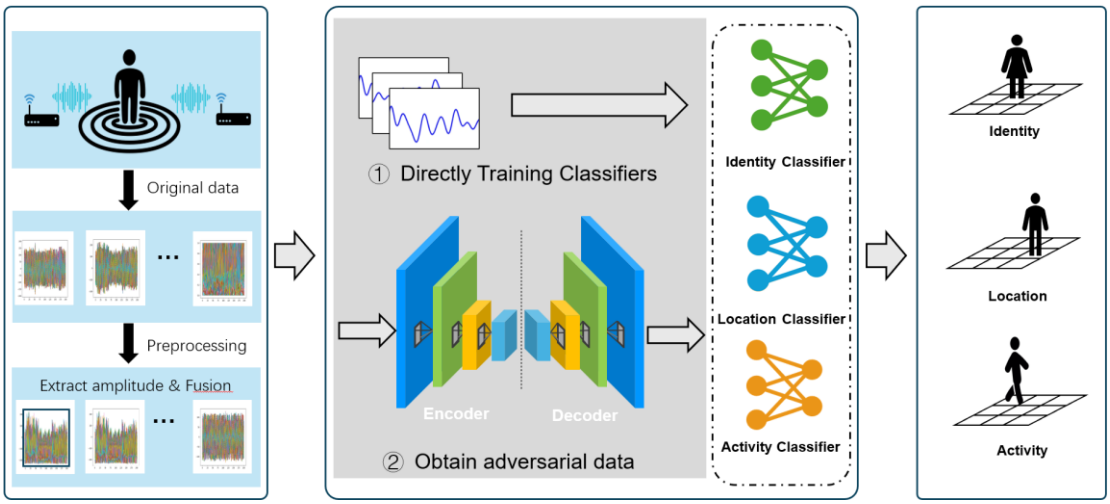



Figure 2: Privacy Protection Process Based on WiFi Sensing

- **Privacy Protection Category in Human Behavior Recognition**

Zhang et al., (2020) classified the privacy protection problem into two types: nontargeted protection and targeted protection. Let $x \in X$ denote the original data, $y^a \in Y^a$, $y^i \in Y^i$, $y^l \in Y^l$, denote the activity, identity, and location labels, respectively. The nontargeted protection can be shown by the Eq. (5) when identity is needed to be protected:

$$a(x') = y^a, l(x') = y^l, \quad s.t. \quad i(x') = y^i(x') \neq y^i, \tag{5}$$

where $y^i(x') \in Y^i$, $a$, $l$, i represent activity classifier, location classifier and identity classifier.$x'$ is adversarial signal.

Targeted protection can be expressed by Eq. (6):

$$a(x') = y^a, l(x') = y^l, \quad s.t. \quad i(x') = y_t^i(x'), \tag{6}$$

where $y_t^i(x') \in Y^i$ denotes the label $t$ predicted by the classifier $i$ with $x'$.

Eq. 6 ensures that the classification result$y_t^i(x')$of the adversarial sample is always the targeted label $t$. In this work, we focus on target protection.

### 2) The Proposed Algorithm ThAN

In this work, we combine 3D-CNN, Attention mechanism and Autoencoder Neural Network to achieve privacy protection while maintaining high performance.

Figure 2 illustrates the process of privacy protection based on CSI signal. Initially, the acquired raw signals undergo a series of preprocessing steps including amplitude extraction, min-max normalization, and data frame fusion, resulting in the Channel State Information (CSI) signals that serve as input data.

Subsequently, these preprocessed CSI signals are fed into three distinct classifiers: an action classifier, an identity classifier, and a location classifier. Through training, these classifiers are optimized to produce accurate model parameters and recognition outcomes. In the third step, the parameters of these classifiers are fixed, and the preprocessed CSI signals are fed into an autoencoder neural network to generate adversarial signals. After multiple training iterations, the desired adversarial signals are produced. These signals are then transmitted to a server for application.

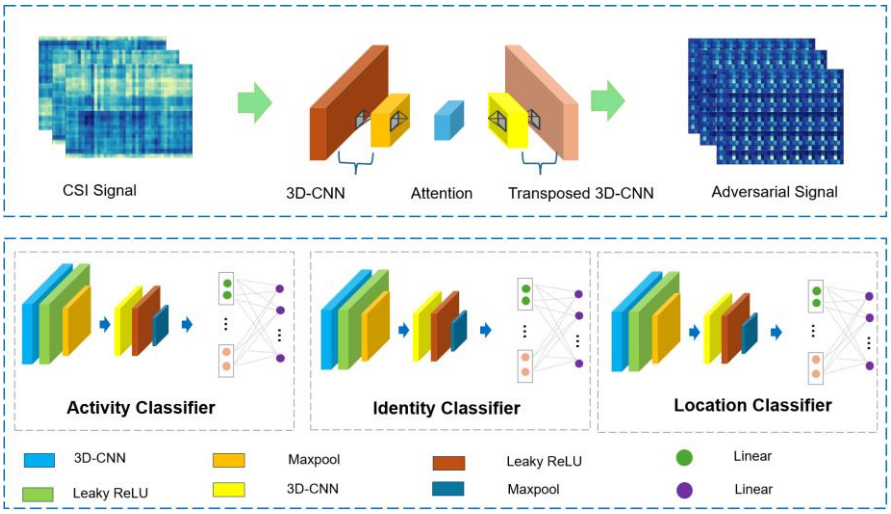



Figure 3: Architecture of the Proposed Algorithm ThAN

Figure 3 explains the architecture of the proposed algorithm. As shown in this figure, the model structures of the action classifier, identity classifier, and location classifier are identical, which facilitates the avoidance of complex network parameter tuning. As aforementioned, 3D-CNNs are effective in extracting features across temporal and spatial dimensions. Consequently, the classifiers employ two layers of 3D-CNNs for feature extraction. For the activity classifier, the features are transformed into a one-dimensional vector consisting of $n_a(\in Z^+)$ features, where na represents the number of action categories. Similarly, for the identity classifier, features are converted into a one-dimensional vector of$n_{\mathrm{i}}(\in Z^+)$features, where ni denotes the number of identity categories. For the location classifier, features are transformed into a one-dimensional vector comprising $n_{\mathrm{l}}(\in Z^+)$features, where nl indicates the number of location categories. Finally, the cross-entropy function Eq. (7) is employed to compute the discrepancy between the predicted values and the true labels.

$$loss = -\sum_{s=1}^{S} y \log f\,(\theta, x), \tag{7}$$

Where $S$ represents the total number of samples, $y_i$ denotes the true labels, $f$ is the Softmax function and x is the one-dimensional vector input to the prediction layer.

After training the three classifiers, an autoencoder neural network model based on 3D-CNN and attention mechanism is designed. By transforming CSI signals into adversarial signals, it is possible to achieve the objective of protecting privacy. The autoencoder neural network model comprises three parts. Initially, a 2-layer 3D-CNN neural network extracts temporal and spatial features; the 3D-CNN is capable of capturing features across both time and space dimensions, making it suitable for processing sequential data. Using multilayer 3D-CNNs can progressively extract higher-level abstract features.

Traditional WiFi-based human behavior recognition algorithms often employ 2D-CNN networks, which cannot extract temporal and spatial features.

Subsequently, an attention module is used to automatically learn and selectively focus on important information within the features. The attention mechanism can automatically learn the importance of different features, thereby selectively focusing on key features to enhance model performance. Traditional autoencoder networks reconstruct the input directly through the encoding-decoding process without explicitly selecting key features.

The generated adversarial signals are input into the three previously trained classifiers to assess the classification differences between the adversarial and real data. During this process, only the autoencoder neural network undergoes learning, while the classifiers do not engage in further learning. The loss function of the autoencoder neural network consists of three parts Eq. 8:

$$loss = loss_a + loss_i + loss_l, \tag{8}$$

Where $loss_a$, $loss_i$, $loss_l$, represent the loss function of the action classifier, identity classifier, and location classifier, respectively.

When identity information needs to be protected, the identity labels of all adversarial signals are set to the identity of the same label $t$, according to to Eq. (6). Finally, the adversarial signals generated during the training process contain only the identity information of the specified user t, thereby protecting the information of other users. Since action and location information need to remain public, there is no need to modify the action and location labels. Hence, the loss function of the autoencoder neural network module is rewritten as Eq. 9:

$$loss = -\left(\sum_{s=1}^{S} y^a \log f_a(\theta, x) + \sum_{s=1}^{S} y_t^i \log f_i(\theta, x) + \sum_{s=1}^{S} y^l \log f_l(\theta, x)\right), \tag{9}$$

Setting the identity labels of all generated adversarial signals to the identity of a single user $t$ effectively protects the identity information of other users; however, it results in the leakage of user $t$'s identity information. To address this issue, we add fictitious blank data to the collected dataset, where each value is set to zero and contains no useful information. Then, the blank data is assigned to the identity label of user $t$, ensuring that even if the user $t$'s identity information is leaked, the leaked information is fictitious, thereby effectively achieving privacy protection.

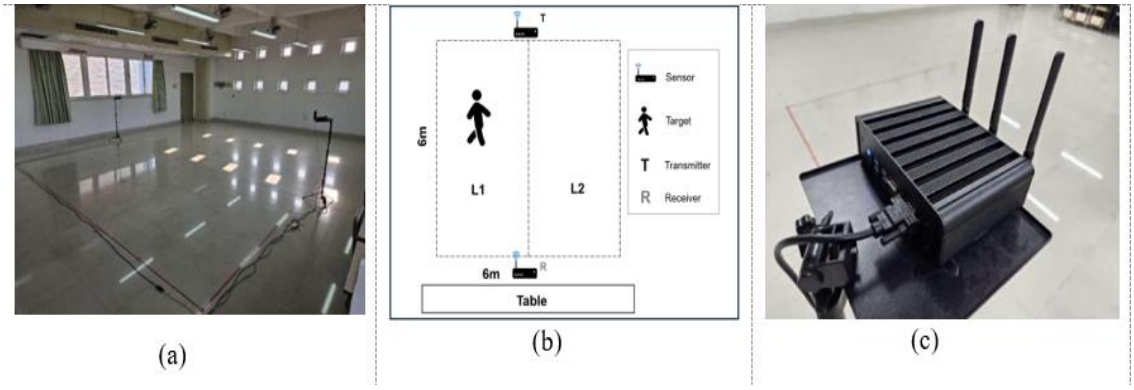



Figure 4: Data Collection Environment

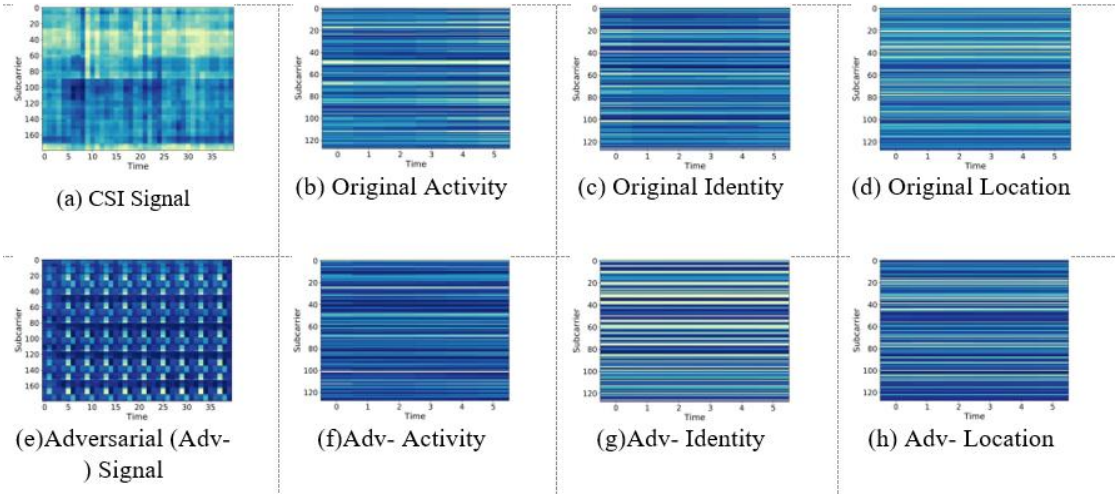



Figure 5: Visualization of Original Signal and Adversarial Signal

# 4 Experimental Evaluation

This study uses a dataset collected in a single room. All experiments were conducted on a workstation equipped with a GeForce RTX 3050 GPU and 16GB of memory, utilizing the TensorFlow 1.18.0 framework and Python 3.6.

### 1) Data Collection and Preprocessing

The data used in this work is collected in a 6m × 6m area as shown in Figure 4. A total of 8 volunteers perform activities; Each activity was performed 30 times resulting in a total of 1200 activity samples. As illustrated above, we add 5 blank samples into our dataset, so the sample size is 1205. The transmitter and receiver were equipped with Qualcomm AR9380. There are 3 transmit antennas and 3 receiver antennas. The number of subcarriers is 56.

To reduce data redundancy and improve training speed, we selected 20subcarriers and utilized keyframe extraction (Huang et al., 2024). As a result, the samples dimensions are 9 × 20 × 40 and each sample is assigned corresponding labels.

### 2) Performance Evaluation of ThAN

In addition to measuring recognition accuracy, we also use the privacy protection success rate (SR) to evaluate the effectiveness of privacy protection, which is calculated by Eq. (10).

$$SR = \frac{T}{S} \times 100\%, \tag{10}$$

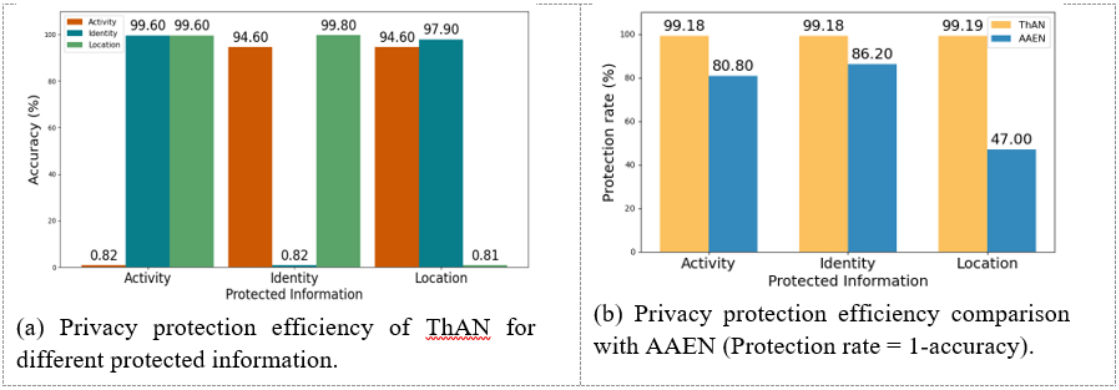
Where $T$ represents the number of adversarial signals predicted as label $t$, and $S$ represents the total number of adversarial signals.

We compare the proposed ThAN with the state-of-the-art AAEN method (Abolqasem et al., 2015). AAEN utilizes a 9-layer 2D-CNN to construct the autoencoder neural network. Science no pooling is applied, the size of the features in each layer remains the same as the input size. The structures of the activity, identity and location classifiers are identical, with each classifier consisting of 2 2D-CNN layers and 2 fully connected layers. Other configurations are shown in Table 1.

Table 1: Network Configuration

| | Classifiers | Autoencoder |
|---|---|---|
| Kernal size | (3,3,3) | (3,3,3) |
| Channel | 32,64 | 16,32,16,1 |
| Learning rate | 0.0001 | 0.0001 |
| Epoch | 10000 | 20000 |
| Padding | Same | Same |
| Batch size | 128 | 128 |

Figure 5 visualizes the original CSI signal and adversarial signal when the protected information is identity. Figure 5(g) is the adversarial signal generated by the autoencoder neural network. It can be observed that there is a significant difference with the original signal Figure 5(c). This indicates that the identity identity-related features within the adversarial signal are successfully modified, substantially reducing the accuracy of identity recognition. As a result, the model's adversarial output can effectively prevent the leakage of identity information. In contrast, the original signal 5(b), 5(d) is very similar to adversarial signal 5(f),5(h), meaning that the adversarial signal preserves relevant activity and location features.




(a) Privacy protection efficiency of ThAN for different protected information.

(b) Privacy protection efficiency comparison with AAEN (Protection rate = 1-accuracy).

Figure 6: Protection Capability of the Proposed ThAN

For protected information, the lower the recognition accuracy, the better; for other information, the higher the recognition accuracy, the better.

Table 2: Recognition Accuracy on Original Data

| Model | Activity | Identity | Location |
|---|---|---|---|
| ThAN | **98.2%** | **99.8%** | **99.8%** |
| AAEN | 96.4% | 98.3% | 99.3% |

From Table 2, it can be observed that when using original data, the recognition accuracy of our proposed method surpasses that of AAEN, with improvements of 1.8%, 1.5%, and 0.5% for action, identity, and location, respectively, demonstrating the superiority of our proposed algorithm. To illustrate privacy protection power, Table 3 shows the comparison result between our proposed ThAN and the AAEN on adversarial data. Regardless of whether the protected information is activity, identity, or location, our proposed method achieves a 99.9% privacy protection success rate, which is 0.4% and 0.8% higher than AAEN when the protected information is identity and location, respectively. As depicted in Figure 6(a), when the protected information is activity, our method achieves an activity recognition accuracy of 0.82%, while the recognition accuracy for unprotected information remains above 99%. In contrast, AAEN still achieves a protected activity recognition accuracy of 19.2%, which means that there is still a one-fifth chance of identifying the true label. The same situation is observed

when the protected information is identity and location as shown in Figure 6(b). The largest gap occurs when the protected information is location, where AAEN still achieves a 53% recognition accuracy for the protected location information, while our proposed method achieves only 0.81%. This indicates that the comparative method still poses a privacy leakage risk of over fifty percent chance, particularly when the protected information is location-based. Furthermore, the recognition accuracy for unprotected information is higher than that of AAEN observed from Table 3.

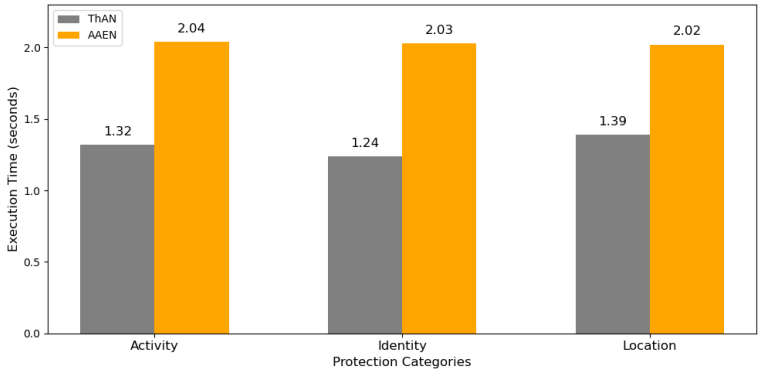



Figure 7: Execution Time Comparison of ThAN and AAEN Across Different Protection Categories

Table 3: Comparison between ThAN and AAEN on Adversarial Data

| Model | Protected Information | SR | Activity | Identity | Location |
|---|---|---|---|---|---|
| ThAN | Activity | 99.9% | **0.82%** | 99.6% | 99.6% |
| | Identity | 99.9% | 94.6% | **0.82%** | 99.8% |
| | Location | 99.9% | 94.6% | 97.9% | **0.81%** |
| AAEN | Activity | 99.9% | **19.2%** | 96.3% | 99.6% |
| | Identity | 99.5% | 90.4% | **13.8%** | 99.2% |
| | Location | 99.1% | 89.8% | 93.3% | **53.0%** |

In addition, From Figure 7, it is evident that ThAN consistently achieves significantly lower execution times across all protection scenarios (Activity, Identity, Location) compared to AAEN. For instance, ThAN reduces the completion time by approximately 35% on Activity (1.32s vs. 2.04s), 39% on Identity (1.24s vs. 2.03s), and 31% on Location (1.39s vs. 2.02s). This improved efficiency and responsiveness indicates that ThAN offers a clear performance advantage over AAEN. Overall, our proposed approach outperforms the AAEN.

## 5 Conclusion

In this work, we proposed a privacy-preserving method for human behavior recognition under AIoT, utilizing 3D-CNNs, transposed 3D-CNNs, and attention modules to construct an autoencoder neural network for generating adversarial signals. Simultaneously, we constructed three classifiers using 3D-CNNs to recognize adversarial samples generated by the autoencoder neural network. This study used the real dataset for the experiment evaluation, to which virtual blank samples were added. Experimental results indicate that the proposed method achieves effective privacy protection, with nearly 99% protection effectiveness for action, identity, and location information while also effectively reducing the impact on the accuracy of recognizing other information.

**Funding:** This work was supported in part by the Guangdong Basic and Applied Basic Research Foundation (Project No. 2021A1515110455), in part by the National Natural Science Foundation of China under Grant 62001126, and in part by the JSPS Grant-in-Aid for Scientific Research (C) 23K11103.

## References

[1] Abanto-Leon, L. F., Bäuml, A., Sim, G. H., Hollick, M., & Asadi, A. (2020). Stay connected, leave no trace: Enhancing security and privacy in wifi via obfuscating radiometric fingerprints. *Proceedings of the ACM on Measurement and Analysis of Computing Systems, 4*(3), 1-31. https://doi.org/10.1145/3428329

[2] Abolqasem, S. S., Sabzevari, S. A., & Kamel, S. R (2015). Developing a routing protocol for wireless sensor networks using fuzzy logic and focused on optimal route election. *International Academic Journal of Science and Engineering, 2*(2), 153–163.

[3] Ahmed, S., & Cho, S. H. (2023). Machine learning for healthcare radars: Recent progresses in human vital sign measurement and activity recognition. *IEEE Communications Surveys & Tutorials*. https://doi.org/10.1109/COMST.2023.3334269

[4] Atawneh, S., Alshammari, Z., Mousa, A. L., & Shawar, B. A. (2024). A Security Framework for Addressing Privacy Issues in the Zoom Conference System. https://doi.org/10.58346/JISIS.2024.I1.016

[5] Awadzi, C., Calloway, J. A., & Awadzi, W. A. (2018). Increasing Conversions through Behavioral Retargeting. *International Academic Journal of Social Sciences, 5*(1), 23–27. https://doi.org/10.9756/IAJSS/V5I1/1810003

[6] Awaisi, K. S., Ye, Q., & Sampalli, S. (2024). A Survey of Industrial AIoT: Opportunities, Challenges, and Directions. *IEEE Access*. https://doi.org/10.1109/ACCESS.2024.3426279

[7] Chithra Devi, S. A., Mahendravarman, I., Ragavendiran, A., Selvam, M. M., Yamuna, K., & Vibin, R. (2024, July). Optimal Configuration of Radial Distribution Networks with Stud Krill Herd Optimization. In *2024 International Conference on Signal Processing, Computation, Electronics, Power and Telecommunication (IConSCEPT)* (pp. 1-6). IEEE. https://doi.org/10.1109/IConSCEPT61884.2024.10627797

[8] Clavijo-López, R., Navarrete, W. A. L., Velásquez, J. M., Saldaña, C. M. A., Ocas, A. M., & Tananta, C. A. F. (2024). Integrating Novel Machine Learning for Big Data Analytics and IoT Technology in Intelligent Database Management Systems. *Journal of Internet Services and Information Security, 14*(1), 206-218. https://doi.org/10.58346/JISIS.2024.I1.014

[9] Dang, X., Liu, Y., Hao, Z., Tang, X., & Shao, C. (2020). Air gesture recognition using WLAN physical layer information. *Wireless Communications and Mobile Computing, 2020*(1), 8546237. https://doi.org/10.1155/2020/8546237

[10] Ding, S., Zhao, P., Zhang, X., Qian, R., Xiong, H., & Tian, Q. (2023). Prune spatio-temporal tokens by semantic-aware temporal accumulation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 16945-16956).

[11] Duan, P., Li, C., Li, J., Chen, X., Wang, C., & Wang, E. (2022). WISDOM: Wi-Fi-Based Contactless Multiuser Activity Recognition. *IEEE Internet of Things Journal, 10*(2), 1876-1886. https://doi.org/10.1109/JIOT.2022.3210131

[12] Fridani, F., & Vosoughi, S. (2014). Morally correct pattern of behavior and social interaction with a focus on Islamic education in the Prophet of Islam (PBUH). *International Academic Journal of Humanities, 1*(1), 60–68.

[13] Ghazi, A., Aljunid, S. A., Idrus, S. Z. S., Fareed, A., Al-dawoodi, A., & Mohsin, A. H. (2021, February). Design of a hybrid WDMA-Optical-CDMA over multi-mode fiber transmission system based on LG modes for short haul-local area network. In *Journal of Physics: Conference*

*Series* (Vol. 1793, No. 1, p. 012016). IOP Publishing. https://doi.org/10.1088/1742-6596/1793/1/012016

[14] Gupta, N., Gupta, S. K., Pathak, R. K., Jain, V., Rashidi, P., & Suri, J. S. (2022). Human activity recognition in artificial intelligence framework: a narrative review. *Artificial intelligence review*, *55*(6), 4755-4808. https://doi.org/10.1007/s10462-021-10116-x

[15] Huang, H., Huang, T., Wang, W., Zhao, L., Wang, H., & Wu, H. (2023). Federated Learning and Convex Hull Enhancement for Privacy Preserving WiFi-Based Device-Free Localization. *IEEE Transactions on Consumer Electronics*. https://doi.org/10.1109/TCE.2023.3342834

[16] Huang, H., Lin, L., Zhao, L., Huang, H., & Ding, S. (2024). TSHNN: Temporal-Spatial Hybrid Neural Network for Cognitive Wireless Human Activity Recognition. *IEEE Transactions on Cognitive Communications and Networking*. https://doi.org/10.1109/TCCN.2024.3414390

[17] Hussain, Z., Sheng, Q. Z., & Zhang, W. E. (2020). A review and categorization of techniques on device-free human activity recognition. *Journal of Network and Computer Applications*, *167*, 102738. https://doi.org/10.1016/j.jnca.2020.102738

[18] Islam, M. M., Nooruddin, S., Karray, F., & Muhammad, G. (2023). Multi-level feature fusion for multimodal human activity recognition in Internet of Healthcare Things. *Information Fusion*, *94*, 17-31. https://doi.org/10.1016/j.inffus.2023.01.015

[19] Jeyanthi, S., & Krishnamoorthi, K. (2022). Quasi Z-source network-based photovoltaic supported STATCOM for voltage and frequency control of stand-alone WECS. *Journal of Circuits, Systems and Computers*, *31*(01), 2250003. https://doi.org/10.1142/S0218126622500037

[20] Ji, S., Xu, W., Yang, M., & Yu, K. (2012). 3D convolutional neural networks for human action recognition. *IEEE transactions on pattern analysis and machine intelligence*, *35*(1), 221-231. https://doi.org/10.1109/TPAMI.2012.59

[21] Jiao, X., Mehari, M., Liu, W., Aslam, M., & Moerman, I. (2021, June). Openwifi CSI fuzzer for authorized sensing and covert channels. In *Proceedings of the 14th ACM Conference on Security and Privacy in Wireless and Mobile Networks* (pp. 377-379). https://doi.org/10.1145/3448300.3468255

[22] Kavitha, M. (2024). Enhancing security and privacy in reconfigurable computing: Challenges and methods. *SCCTS Transactions on Reconfigurable Computing, 1*(1), 16-20. https://doi.org/10.31838/RCC/01.01.04

[23] Kavitha, M. (2024). Environmental monitoring using IoT-based wireless sensor networks: A case study. *Journal of Wireless Sensor Networks and IoT, 1*(1), 32–36.

[24] Klavin, C. (2024). Analysing Antennas with Artificial Electromagnetic Structures for Advanced Performance in Communication System Architectures. *National Journal of Antennas and Propagation*, *6*(1), 23-30. https://doi.org/10.31838/NJAP/06.01.04

[25] Kumar, T. M. S. (2024). Security challenges and solutions in RF-based IoT networks: A comprehensive review. *SCCTS Journal of Embedded Systems Design and Applications, 1*(1), 19-24. https://doi.org/10.31838/ESA/01.01.04

[26] Li, Y., Yang, G., Su, Z., Li, S., & Wang, Y. (2023). Human activity recognition based on multi environment sensor data. *Information Fusion*, *91*, 47-63.

[27] Liang, W., Tang, R., Jiang, S., Wang, R., Zhao, Y., Xu, C. Z., ... & Li, X. (2023). LiWi-HAR: Lightweight WiFi-Based Human Activity Recognition Using Distributed AIoT. *IEEE Internet of Things Journal*, *11*(1), 597-611. https://doi.org/10.1109/JIOT.2023.3286455

[28] Luo, F., Bodanese, E., Khan, S., & Wu, K. (2023). Spectro-temporal modeling for human activity recognition using a radar sensor network. *IEEE Transactions on Geoscience and Remote Sensing*, *61*, 1-13. https://doi.org/10.1109/TGRS.2023.3270365

[29] Meng, W., Chen, X., Cui, W., & Guo, J. (2021). WiHGR: A robust WiFi-based human gesture recognition system via sparse recovery and modified attention-based BGRU. *IEEE Internet of Things Journal*, *9*(12), 10272-10282. https://doi.org/10.1109/JIOT.2021.3122435

[30] Meng, W., Liu, Z., Li, B., Cui, W., Zhou, J. T., & Zhang, L. (2023). GrapHAR: A Lightweight Human Activity Recognition Model by Exploring the Sub-carrier Correlations. *IEEE Transactions on Wireless Communications*. https://doi.org/10.1109/TWC.2023.3302861

[31] Miao, F., Huang, Y., Lu, Z., Ohtsuki, T., Gui, G., & Sari, H. (2024). Wi-Fi Sensing Techniques for Human Activity Recognition: Brief Survey, Potential Challenges, and Research Directions. *ACM Computing Surveys*. https://doi.org/10.1145/3705893

[32] Miki, M., Yamauchi, T., & Kobayashi, S. (2024). Effectiveness of MAC Systems based on LSM and their Security Policy Configuration for Protecting IoT Devices. *Journal of Internet Services and Information Security, 14*(3), 293-315. https://doi.org/10.58346/JISIS.2024.I3.018

[33] Moghaddam, M. G., Shirehjini, A. A. N., & Shirmohammadi, S. (2023). A WiFi-based method for recognizing fine-grained multiple-subject human activities. *IEEE Transactions on Instrumentation and Measurement*, *72*, 1-13. https://doi.org/10.1109/TIM.2023.3289547

[34] Muaaz, M., Chelli, A., Gerdes, M. W., & Pätzold, M. (2022). Wi-Sense: A passive human activity recognition system using Wi-Fi and convolutional neural network and its integration in health information systems. *Annals of Telecommunications*, *77*(3), 163-175.

[35] Nwosu, P. O., & Adeloye, F. C. (2023). Transformation leader strategies for successful digital adaptation. *Global Perspectives in Management, 1*(1), 1–16.

[36] Pan, Y., Zhou, Z., Gong, W., & Fang, Y. (2024). SAT: A Selective Adversarial Training Approach for WiFi-based Human Activity Recognition. *IEEE Transactions on Mobile Computing*. https://doi.org/10.1109/TMC.2024.3420405

[37] Rezaei, A., Stevens, M. C., Argha, A., Mascheroni, A., Puiatti, A., & Lovell, N. H. (2022). An unobtrusive human activity recognition system using low resolution thermal sensors, machine and deep learning. *IEEE Transactions on Biomedical Engineering*, *70*(1), 115-124. https://doi.org/10.1109/TBME.2022.3186313

[38] Shinan, K., Alanazi, F., Alhazmi, H. E., Riaz, F., Fatima, N., Abid, A., ... & Ashraf, M. U. (2024). EUAC: An Advanced Privacy Protection for Location-based Service System. *Journal of Internet Services and Information Security*, *14*(2), 202-225. https://doi.org/10.58346/JISIS.2024.I2.013

[39] Wang, D., Yang, J., Cui, W., Xie, L., & Sun, S. (2022). CAUTION: A Robust WiFi-based human authentication system via few-shot open-set recognition. *IEEE Internet of Things Journal*, *9*(18), 17323-17333. https://doi.org/10.1109/JIOT.2022.3156099

[40] Wang, P., Huang, H., Zhao, L., Zhu, B., Huang, H., & Wu, H. (2024). ExtRe: Extended Temporal-Spatial Network for Consumer-Electronic WiFi-Based Human Activity Recognition. *IEEE Transactions on Consumer Electronics*. https://doi.org/10.1109/TCE.2024.3435881

[41] Xianjia, M. E. N. G., Lin, F. E. N. G., Hao, C. H. E. N., Ting, C. H. E. N., Jianfeng, M. A., Anwen, W. A. N. G., ... & Yanfeng, Z. H. A. O. (2021). Just-in-Time Human Gesture Recognition Using WiFi Signals. *Chinese Journal of Electronics*, *30*(6), 1111-1119. https://doi.org/10.1049/cje.2021.07.022

[42] Yadav, S. K., Sai, S., Gundewar, A., Rathore, H., Tiwari, K., Pandey, H. M., & Mathur, M. (2022). CSITime: Privacy-preserving human activity recognition using WiFi channel state information. *Neural Networks*, *146*, 11-21. https://doi.org/10.1016/j.neunet.2021.11.011

[43] Zhang, W., Zhou, S., Peng, D., Yang, L., Li, F., & Yin, H. (2020). Understanding and modeling of WiFi signal-based indoor privacy protection. *IEEE Internet of Things Journal*, *8*(3), 2000-2010. https://doi.org/10.1109/JIOT.2020.3015994

[44] Zhang, X., He, F., Chen, Q., Jiang, X., Bao, J., Ren, T., & Du, X. (2022). A differentially private indoor localization scheme with fusion of WiFi and bluetooth fingerprints in edge computing. *Neural Computing and Applications*, *34*(6), 4111-4132. https://doi.org/10.1007/s00521-021-06815-9

[45] Zhang, Y., Liu, Q., Wang, Y., & Yu, G. (2022). CSI-based location-independent human activity recognition using feature fusion. *IEEE Transactions on Instrumentation and Measurement*, *71*, 1-12. https://doi.org/10.1109/TIM.2022.3216419
[46] Zhao, L., Huang, H., Wang, W., & Zheng, Z. (2023). An accurate approach of device-free localization with attention empowered residual network. *Applied Soft Computing*, *137*, 110164. https://doi.org/10.1016/j.asoc.2023.110164
[47] Zhao, L., Yang, Q., Huang, H., Guo, L., & Jiang, S. (2024). Intelligent wireless sensing driven metaverse: A survey. *Computer Communications*, *214*, 46-56. https://doi.org/10.1016/j.comcom.2023.11.024
[48] Zhao, P., Liu, W., Zhang, G., Li, Z., & Wang, L. (2020). Preserving privacy in WiFi localization with plausible dummy locations. *IEEE Transactions on Vehicular Technology*, *69*(10), 11909-11925. https://doi.org/10.1109/TVT.2020.3006363
[49] Zheng, Y., Zhang, Y., Qian, K., Zhang, G., Liu, Y., Wu, C., & Yang, Z. (2019, June). Zero-effort cross-domain gesture recognition with Wi-Fi. In *Proceedings of the 17th annual international conference on mobile systems, applications, and services* (pp. 313-325). https://doi.org/10.1145/3307334.3326081
[50] Zhou, S., Zhang, W., Peng, D., Liu, Y., Liao, X., & Jiang, H. (2019). Adversarial WiFi sensing for privacy preservation of human behaviors. *IEEE Communications Letters*, *24*(2), 259-263. https://doi.org/10.1109/LCOMM.2019.2952844

## Authors Biography

**Haoda Wang** received his Bachelor's degree from the Central South University of Forestry and Technology, China, in 2017, and acquired his Master's degree from the Monash University, Australia, in 2021. Currently, he is a doctoral student in the School of Computer Science and Engineering at the University of Aizu, Japan. His research interests focus on the Federated Learning and Privacy Preserving.

**Liang Lin** received his B.S. and his M.S. degree from Guangzhou University, China, in 2019 and 2024, respectively. He is currently an assistant professor with the Luoding Polytechnic, China. His current research interests include wireless gesture recognition and deep learning.

**Huakun Huang** (Member, IEEE) received the Ph.D. degree in computer science and engineering from the University of Aizu, Japan, in 2019. He was a Research Fellow with the Aizu Computer Science Laboratories, Inc. from September 2019 to September 2020, and a Visiting Research Fellow with the School of Computer Science and Engineering, University of Aizu from September 2019 to November 2020. He is currently an Associate Professor with the School of Computer Science and Cyber Engineering, Guangzhou University, China. His current research interests include intelligent wireless sensing technology, deep learning, and sparse coding. He has served as a Lead Guest Editor for the Computer Communications.

**Lingjun Zhao** received her Ph.D. in the School of Computer Science and Engineering from the University of Aizu, Japan, in 2019. From October 2019 to October 2020, she was on a postdoctoral fellowship with the Department of Electronic and Information Engineering, the Hong Kong Polytechnic University. She was a postdoctoral fellowship with the School of Software Engineering, at Sun Yat-Sen University, China. She is now an Associate Professor in the School of Electronic and Information, at Guangdong Polytechnic Normal University. Her current research interests include deep learning, privacy preserving technologies, and wireless sensing technology.

**Zhuotao Lian** received his B.S. in Computer Science from China University of Geosciences, Wuhan, in 2020, and his M.S. and Ph.D. in Computer Science and Engineering from the University of Aizu, Japan, in 2021 and 2024, respectively. He was awarded the NEC C&C Foundation Grants for Researchers in 2023 and the Grant-in-Aid for JSPS Fellows in 2024. Currently, he is a JSPS International Research Fellow at Kyushu University. His research interests include federated learning, differential privacy, blockchain technologies, and AI.

**Chunhua Su** received the B.S. degree for Beijing Electronic and Science Institute in 2003 and received his M.S. and PhD of computer science from Faculty of Engineering, Kyushu University in 2006 and 2009, respectively. He is currently working as a Senior Associate Professor in Division of Computer Science, University of Aizu. He has worked as a postdoctoral fellow in Singapore Management University from 2009-2011 and a research scientist in Cryptography & Security Department of the Institute for Infocomm Research, Singapore from 2011-2013. From 2013-2016, he has worked as an Assistant professor in School of Information Science, Japan Advanced Institute of Science and Technology. From 2016-2017, he worked as Assistant Professor in Graduate School of Engineering, Osaka University. His research interests include cryptanalysis, cryptographic protocols, privacy-preserving technologies in machine learning and IoT security & privacy. He has published more than 100 papers in international journals and conferences.