Computers, Materials & Continua

Tech Science Press

Doi:10.32604/cmc.2025.061869

ARTICLE

A Lightweight Convolutional Neural Network with Squeeze and Excitation Module for Security Authentication Using Wireless Channel

Xiaoying Qiu[1,*], Xiaoyu Ma[1], Guangxu Zhao[1], Jinwei Yu[2], Wenbao Jiang[1], Zhaozhong Guo[1] and Maozhi Xu[3]

[1]College of Computer Science, Beijing Information Science and Technology University, Beijing, 100192, China

[2]School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, 100876, China

[3]School of Mathematical Sciences, Peking University, Beijing, 100871, China

*Corresponding Author: Xiaoying Qiu. Email: 20192329@bistu.edu.cn

Received: 05 December 2024; Accepted: 10 February 2025; Published: 16 April 2025

ABSTRACT: Physical layer authentication (PLA) in the context of the Internet of Things (IoT) has gained significant attention. Compared with traditional encryption and blockchain technologies, PLA provides a more computationally efficient alternative to exploiting the properties of the wireless medium itself. Some existing PLA solutions rely on static mechanisms, which are insufficient to address the authentication challenges in fifth generation (5G) and beyond wireless networks. Additionally, with the massive increase in mobile device access, the communication security of the IoT is vulnerable to spoofing attacks. To overcome the above challenges, this paper proposes a lightweight deep convolutional neural network (CNN) equipped with squeeze and excitation module (SE module) in dynamic wireless environments, namely SE-ConvNet. To be more specific, a convolution factorization is developed to reduce the complexity of PLA models based on deep learning. Moreover, an SE module is designed in the deep CNN to enhance useful features and maximize authentication accuracy. Compared with the existing solutions, the proposed SE-ConvNet enabled PLA scheme performs excellently in mobile and time-varying wireless environments while maintaining lower computational complexity.

KEYWORDS: Physical layer authentication; blockchain; squeeze and excitation module; computational cost; mobile scenario

1 Introduction

Advancements in wireless technology are driving the rapid deployment of Internet of Things (IoT) devices, enabling extensive connectivity across various applications and environments [1]. However, the increasing transmission of sensitive data over wireless channels has raised significant concerns about the security of IoT network communications [2]. Given the massive scale of mobile device access, ensuring IoT communication security has become increasingly critical [3]. Blockchain, a type of distributed ledger technology, is utilized in cryptographic algorithms for its secure, reliable, and tamper-proof characteristics [4]. Fang et al. [5] proposed a smooth handover authentication approach that utilizes decentralized edge intelligence and hierarchical blockchain technology in the zero-trust Internet of Vehicles (IoV). However, blockchain technology inherently relies on the execution of complex cryptographic algorithms and consensus mechanisms, both of which demand significant computational resources [6]. To enhance the utilization of computing resources, researchers have turned to physical layer authentication (PLA) technique that offers distinct advantages over traditional upper-layer authentication methods based on encryption. Unlike encryption, PLA is characterized by low latency and reduced complexity [7]. In light of the security threats posed by potential attackers, a rapid and efficient lightweight authentication scheme is essential for identifying unknown IoT transmitting devices effectively [8].

In recent years, PLA has proven to be a promising approach for detecting spoofing attacks by utilizing physical layer characteristics [9]. PLA differentiates between legitimate and illegitimate devices by analyzing radio frequency (RF) [10] fingerprints, including received signal strength (RSS), and carrier frequency offset (CFO). However, a device's RF fingerprint arises from inherent hardware imperfections and is influenced by environmental conditions, such as temperature. Time-varying factors can also impact these fingerprints, posing challenges for reliable authentication [11]. RSS and CFO have limitations as they only provide limited information on signal strength and frequency offset, and fail to capture the multi-dimensional characteristics of the channel, such as phase and delay [12]. Channel state information (CSI) provides a more detailed and fine-grained fingerprint of the wireless channel. Based on the widely used Jakes model, received signals become decorrelated beyond a distance of half a wavelength, making it challenging for attackers to replicate the channel between legitimate devices [13]. Therefore, CSI fingerprints offer high spatial resolution, making them highly effective for detecting illegitimate devices [14].

CSI provides detailed insights into channel conditions, including amplitude and phase data across multiple channels [15]. Such fine-grained information allows CSI-based authentication methods to distinguish more accurately between devices or devices by leveraging specific channel characteristics. PLA can be broadly classified according to the decision strategy into threshold-based and threshold-free methods. Additionally, it can be further subdivided based on the characteristics of the authenticated devices into categories such as static user authentication and mobile user authentication.

On the one hand, some researchers propose physical layer authentication in static scenarios. Liao et al. [16] and Li et al. [17] used convolutional neural networks (CNNs) to differentiate legitimate and illegitimate devices under simulated conditions, achieving a certain degree of accuracy. In [18], Qiu et al. introduced an adaptive neural network capable of tracking time-varying CSI data for intelligent authentication. In [19], the use of a confidence branch in combination with CNNs enabled not only discrimination between legitimate and illegitimate devices but also identification of distinct legitimate devices. Further, Gao et al. [20] proposed an angle-delay calculation method using environmental semantics, leveraging the YOLO network and a lightweight architecture to extract CSI data and achieve high-precision authentication. Xie et al. [21] suggested utilizing phase differences to map wireless device locations for spoofing attack detection, feeding phase-difference-generated heatmaps into a neural network model for training. The results successfully demonstrated the superior performance of this approach. In [22], Chen et al. propose a Convolutional Denoising Autoencoder (CDAE) structure to denoise CSI with noise, and then used K-Nearest Neighbors (K-NN) algorithm to distinguish between legitimate and illegitimate devices. Chen et al. [23] use a threshold-based detection approach utilizing channel differences to create labeled offline training datasets for machine learning algorithms, eliminating the need for manual labeling. Martins et al. [24] propose using CSI parameters as an authentication mechanism and tracking the wireless channel to prevent network attacks from impersonating legitimate users' communication. On the other hand, several studies explore the implementation of physical layer security authentication in mobile environments. Study [25] present a deep learning-based authentication approach that captures and monitors changes in channel characteristics, improving the adaptability of PLA. In [26], Jing et al. introduce a multi-attribute PLA mechanism without the need for thresholds. Wang et al. [27] introduced a spatiotemporal gradient-based physical layer authentication (STG-PLA) algorithm enhanced by CSI-to-image transformation. It extracts correlation and scattering features to represent channel characteristics and converts multiple CSI sequences into a CSI image for efficient analysis. In [28], Pan et al. apply residual network (ResNet) to achieve legal and illegal residual network authentication in the mobile industrial Internet scene, and proved that the accuracy of ResNet authentication in the mobile scene is higher than the K-NN algorithm. In [29], data augmentation and transfer learning are introduced into the authentication scheme. ResNet50 network is used to realize multi-user authentication in the industrial IoT environment.

However, PLA has been extensively studied, addressing the memory and computational constraints of authentication systems in dynamic, time-varying mobile environments remains a challenging aspect. This paper aims to enhance authentication performance for mobile devices in dynamic environments by introducing a lightweight model for resource-constrained devices, addressing the real-time needs of IoT devices. We collect instantaneous CSI data from the IoT environment to use as training data for our model. When an external device attempts to intrude and submits an access request, its instantaneous CSI data is input into the trained authentication model to extract intricate CSI features, allowing for precise identification of the unknown device's type. Therefore, we introduce the Squeeze and Excitation Convolutional Network (SE-ConvNet), a network that reduces complexity through the integration of the SE module and convolution factorization, while leveraging the squeeze and excitation mechanism of the SE module to enhance the accuracy of the authenticator. Our contributions are summarized below:

1. We propose a scheme based on the SE module and convolution factorization for authenticating CSI from mobile devices in IoT. The goal is to balance authentication accuracy and computation complexity for mobile devices.
2. In order to reduce the computation complexity, we introduce the SE module and convolution factorization. The SE module enhances useful features and suppresses irrelevant ones, demonstrating outstanding performance in various image processing tasks. Convolution factorization decomposes complex convolution operations into simpler ones, reducing computational complexity.
3. Comprehensive evaluation metrics are proposed to evaluate the proposed scheme. Simulation findings indicate that SE-ConvNet not only effectively verifies device identity but also demonstrates robustness in time-varying wireless conditions.

The rest of the paper is as follows: the Section 2 describes our proposed PLA framework in detail through a flowchart, explaining the key methods and principles of its main components. In Section 3, we present the experimental setup, results, and comparisons with alternative algorithms. Finally in the Section 4, we summarizes our findings and discusses potential directions for future research.

2 Methodology

In this section, we begin by presenting the system model for PLA. Next, we detail the proposed SE-ConvNet, including its mechanism and the design of the authentication model. Finally, we provide an analysis of the model's complexity.

2.1 System Model

PLA must account for the time-varying characteristics of wireless channels, including fading and multipath effects. Additionally, it should consider the mobility of both the transmitter and receiver. Device movement can cause fluctuations in channel characteristics, influenced by factors such as Doppler shift and coherence time. Fig. 1 depicts the standard Alice-Bob-Eve model in a wireless communication scenario. In this model, Alice, the legitimate transmitter, communicates with Bob, the legitimate receiver, in a multipath scattering environment. Meanwhile, a malicious device, Eve, attempts to impersonate Alice and gain illegitimate access to Bob. Such an attack diminishes the credibility of the legitimate transmitter, posing a threat to the security of the communication system.

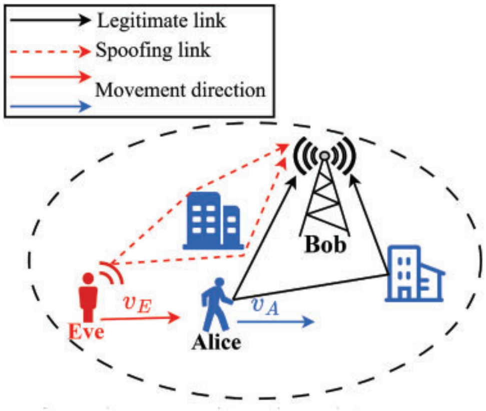

Figure 1: PLA model in the mobile scenario

Assuming that orthogonal frequency division multiplexing (OFDM) technology is utilized in this study and N_s subcarriers are assigned for communication, let the vectors $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$ denote the transmitted and received signals, respectively. We have the following equation:

$$\vec{\mathbf{Y}} = \mathbf{H}_a \cdot \vec{\mathbf{X}} + \mathbf{W} \tag{1}$$

where $\mathbf{H}_a$ represents the channel frequency response (CFR) matrix with dimensions $N_a \times N_s$, where N_a is the number of antennas and N_s is the number of subcarriers. This matrix can be estimated using pilot signals. The term $\mathbf{W}$ represents additive white Gaussian noise. For the i-th antenna and the j-th subcarrier, the CFR $\mathbf{H}_{a_{ij}}$ is defined as a complex number:

$$\mathbf{H}_{a_{ij}} = \left|\mathbf{H}_{a_{ij}}\right| e^{j\angle \mathbf{H}_{a_{ij}}} = \text{real}(\mathbf{H}_{a_{ij}}) + j \cdot \text{img}(\mathbf{H}_{a_{ij}}) \tag{2}$$

The channel matrix $\mathbf{H}_{a_{ij}}$ typically represents the CSI. Changes in the wireless communication environment can affect some signal paths, while others remain invariant. The receiver (Bob) collects the CSI between itself and the transmitter (Alice), and uses this CSI as a reference for identifying legitimate communication. Given that Eve, an eavesdropper within communication range, may attempt to inject illegitimate messages, Bob must determine the legitimacy of a message source based on the real-time CSI characteristics.

To facilitate authentication, Bob labels the CSI data with unique identifiers to distinguish between transmitters. CSI from Eve is labeled as 0, while Alice's data is labeled as 1. Bob uses historical CSI data and corresponding labels to train the lightweight SE-ConvNet network for authentication. When Bob receives a signal from an unidentified transmitter, he inputs the estimated CSI into the pretrained SE-ConvNet model. By activating the sigmoid function at the end of the fully connected layer, a prediction score mapped to between 0 and 1 is obtained. Finally, the score is compared to the threshold:

$$D = \begin{cases} 1, & \text{if score} > \varepsilon_0 \\ 0, & \text{if score} \leq \varepsilon_0 \end{cases} \tag{3}$$

where $D = 1$ denotes that there is a legitimate device, and $D = 0$ denotes that there is no legitimate devices and threshold ε_0 is obtained through experiments.

2.2 Proposed PLA Scheme

2.2.1 Establishment of Authentication Mechanism

The proposed authentication mechanism is shown in Fig. 2. It consists of three stages: acquisition of instantaneous CSI data, extraction of deep CSI features and establishment of the authentication model, and access authentication for unknown devices.

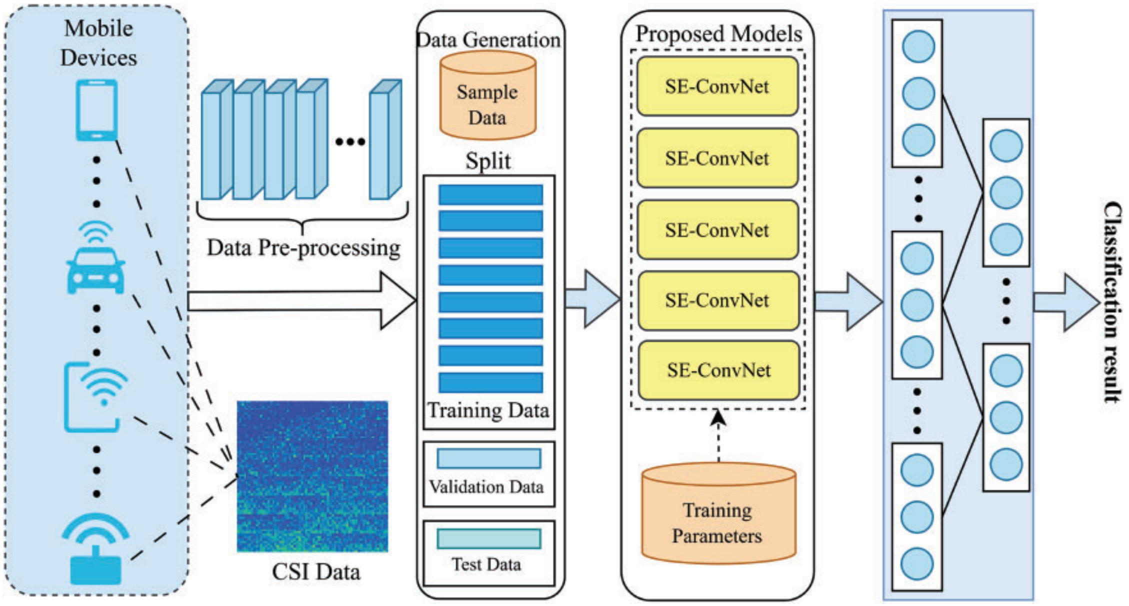

Figure 2: PLA mechanism based on SE-ConvNet

We divide the wireless communication environment into multiple grids, sampled multiple devices moving differently within this space, and collected numerous snapshots of instantaneous CSI data. This approach yields a substantial dataset, ensuring ample training samples for effective model development. Given the distinctiveness of wireless channels, the channel characteristics between each legitimate device and the access point (AP) are unique, which makes them suitable for authentication. When an unknown device attempts to connect to the AP, we capture its real-time CSI data and feed it into the pre-trained authentication model. This process enables the extraction of deep features from the CSI, allowing us to accurately verify the legitimacy of the device's identity. The mathematical expression for this phase is represented in Eq. (4) as follows:

$$C_{\text{class}}(t) = f_{\text{conv}}(\mathbf{H}_a(t)) \tag{4}$$

where $f_{\text{conv}}(\cdot)$ denotes the proposed PLA mechanism, which effectively differentiates between the types of user authentication. The term $\mathbf{H}_a(t)$ represents the current CSI, while $C_{\text{class}}(t)$ indicates the output authentication identity type. The establishment process of the proposed authentication mechanism is illustrated in Algorithm 1.

Algorithm 1: SE-ConvNet physical layer authentication

```
Input: The transmitter's channel $\hat{\mathbf{H}}_a(t)$ in experimental scenario
Output: The updated weight of SE-ConvNet corresponds to the predicted labels for the active transmitters.
Initialize a preprocessed dataset data consisting of the recorded CSI data along with their associated labels.
The pre-processed data set is input to the SE-ConvNet network for training to obtain the output label of the network
if $\hat{y} = 0$ then
    Reject the request and trigger a spoofing alarm.
 else
     Accept the request and identify which legitimate transmitter it originated from.
end if
Return the prediction tag for $K$ devices $\{y_1, \ldots, y_K\}$ and updated weight matrix
```

2.2.2 The Structure of SE-ConvNet

As the core of the proposed PLA mechanism, the design of SE-ConvNet has a great impact on the overall authentication performance. Fig. 3 illustrates the architecture of SE-ConvNet.

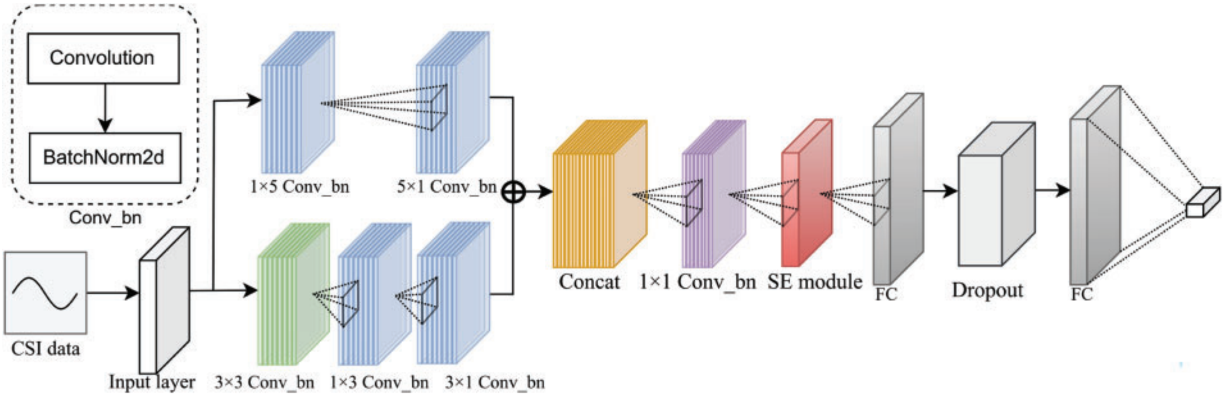

Figure 3: The structure of SE-ConvNet

After each convolution layer in the SE-ConvNet architecture is a Batch Normalization (BN) layer and a Leaky ReLU activation function. Inspired by ResNet [30] and channel attention mechanism [31], we introduce the SE module suitable for lightweight processing while designing the network structure. During network training, the BN layer normalizes a group of sample features with the same dimension, which speeds up model training and improves accuracy. The activation function gives the network nonlinear modeling capabilities, allowing it to perform nonlinear transformations on separable CSI data and enabling the verifier to capture and model underlying core features. SE-ConvNet applies Leaky ReLU to the various convolution layers. The Leaky ReLU formula is shown below:

$$\text{Leaky ReLU}(x) = \begin{cases} x, & \text{if } x \geq 0 \\ \alpha x, & \text{if } x < 0 \end{cases} \tag{5}$$

where $\alpha \in (0,1)$ is the negative slope. We set α to 0.3 in the SE-ConvNet. The various components of the SE-ConvNet network structure are integral to its functionality, and these components are described in detail below.

Convolution Factorization Convolution factorization is a technique that decomposes a standard convolution kernel into multiple smaller kernels. This approach is widely utilized in deep learning, particularly in the design of CNNs. In this paper, convolution factorization is incorporated into the proposed SE-ConvNet architecture to reduce both computational complexity and the number of parameters, while striving to maintain model accuracy as much as possible. The structure is illustrated in Fig. 4.

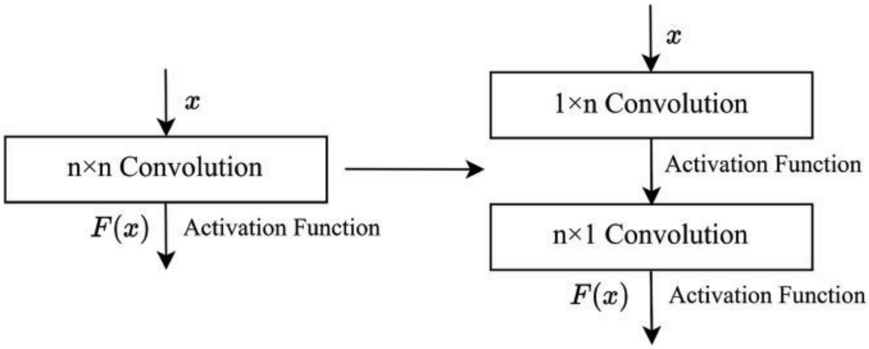

Figure 4: Convolution factorization

This approach involves decomposing the standard convolution kernel into multiple smaller convolution kernels. The core concept of convolution factorization is rooted in tensor decomposition theory, which has been extensively applied in signal processing, pattern recognition, and machine learning. By decomposing high-dimensional tensors into the product of multiple low-dimensional tensors, this approach effectively reduces both storage requirements and computational costs. This principle has been seamlessly incorporated into the optimization of convolution operations in deep learning, particularly in the development of lightweight models and their deployment on mobile and embedded systems. In our proposed design, convolution factorization is integrated into the SE-ConvNet architecture, with the specific calculation formula represented as:

$$U_{\mathscr{N}}^{(l+1)}(i,j) = f_l\left(\sum_{k=1}^{K_l}\left[\sum_{x=1}^{n_l} U_k^{(l)}(s_l \cdot i + x, j) w_{\mathscr{K}_1}^{(l)}(x,y)\right] \cdot \sum_{y=1}^{n_l} w_{\mathscr{K}_2}^{(l)}(y,1) + \Xi\right) \tag{6}$$

where $U^{(l)}$ refers to the output of neurons in layer l, while $f^{(l)}(\cdot)$ represents their activation function. $\mathscr{K}_1$ and $\mathscr{K}_2$ denote convolution kernels of size $1 \times n$ and $n \times 1$, respectively. The dimensions of the neurons' output in layer l are specified as $n_l \times n_l$. K_l and n_l correspond to the number of convolution kernels and the number of neurons in this layer, respectively. The weight of the k-th convolution kernel in layer l is denoted as w_k^l. Parameters s_l and p_l define the stride and the size of padding applied in the convolution layer. The threshold matrix is represented by Ξ. Padding involves surrounding the edges of the input matrix with additional elements (such as zeros, ones, or repeated values) to ensure that the output matrix has the same dimensions as the input after convolution. Stride determines how far the convolution kernel moves with each step during the operation. This decomposition method effectively reduces computational complexity while preserving the model's expressiveness.

SE Module We introduce the SE module, which is designed based on the principles of Squeeze and Excitation Networks. The objective of this module is to enhance model accuracy while maintaining a low floating-point operations (FLOPS) count. The structure of the SE Module is illustrated in Fig. 5.

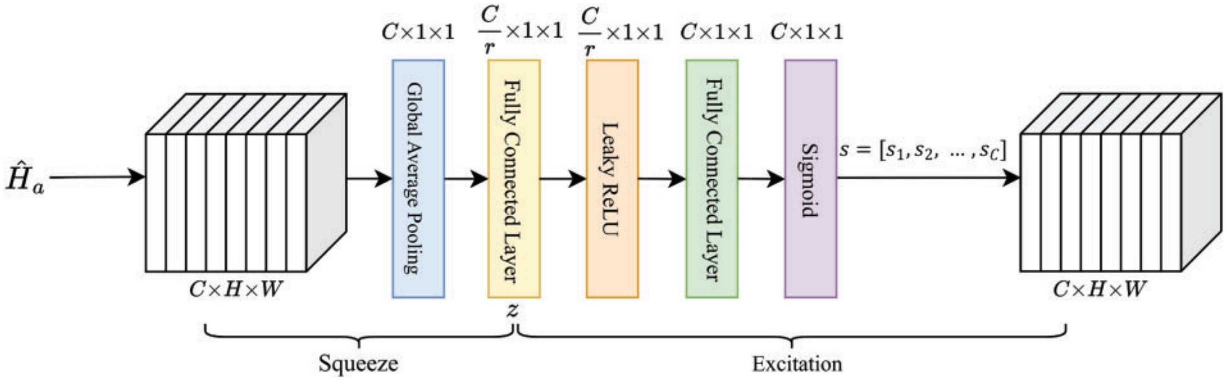

Figure 5: The structure of SE module

Step 1 Squeeze: Each learned filter operates within a local receptive field, which means that each unit of the transformation output global channel information cannot leverage contextual information beyond that localized region, the formula to be applied is $F(\hat{U}(i,j)) = D(i,j)$. This issue becomes more pronounced at the lower levels of the network, where the receptive field is relatively small. To mitigate this limitation, we propose compressing the global channel information into a single channel descriptor. This can be accomplished by employing global average pooling to generate channel-level statistics. Formally, in the spatial dimension of the narrow $h \times w$ output D_c, we generate statistics z, with the first Z_c elements calculated using the following formula:

$$Z_c = \frac{1}{h \times w} \sum_{i=1}^{h} \sum_{j=1}^{w} D_c(i,j), \quad \text{for } c = 1, \dots, C \tag{7}$$

Step 2 Excitation: To maximize the information gathered during the compression process, a second operation is performed to effectively capture channel dependencies. This operation needs to meet two key requirements: firstly, it must be able to learn nonlinear interactions between channels; secondly, it should account for non-mutually exclusive relationships, enabling multiple channels to be considered at the same time rather than isolating a single activation. To fulfill these requirements, we have selected a simple gating mechanism and employed the sigmoid activation function:

$$s = \sigma(W_2 \cdot \delta(W_1 z)) \tag{8}$$

the channel weight s generated by the excitation operation provides crucial information for subsequent weighting. We apply these weights to the original feature map u, enabling the model to dynamically adjust the significance of each feature based on channel importance. Dimensionality is reduced through the fully connected layer with a Leaky ReLU activation function, usually reducing the dimension from C to C/r (where r is the scaling factor). This approach enhances the expressive power and overall performance of the network. The channel weight s is applied to the original feature map D, with weighting achieved through channel-level multiplication. Consequently, the output feature map can be expressed as follows:

$$\widetilde{X} = D \cdot s \tag{9}$$

2.2.3 Complexity Analysis of SE-ConvNet

To some degree, the number of FLOPS reflects the computational complexity of SE-ConvNet. The complexity of a single convolution layer can be expressed as $O(h_l w_l k_l^2 C_{l-1} C_l)$. The convolution factor complexity introduced in this paper is given by $O(h_l w_l (k_l + 1) C_{l-1} C_l)$. Because the SE module utilizes fully connected layers and global average pooling layers, its computational complexity can be described as:

$$O\left(C \cdot \left(\frac{C}{r}\right) + \left(\frac{C}{r}\right) \cdot C\right) = O\left(\frac{2C^2}{r}\right) = O\left(\frac{C^2}{r}\right) \tag{10}$$

By integrating the complexity of the convolution layers with the complexity formula of the SE module, the overall complexity of SE-ConvNet can be derived as follows:

$$O\left(h_l w_l (k_l + 1) C_{l-1} C_l + \frac{C^2}{r}\right) \tag{11}$$

where h_l and w_l denote the height and width of the feature map, with k_x representing the size of the filter. C_{l-1} is the number of output channels from the preceding layer, and C_l refers to the output channels in the current layer. The relationship between the output feature at layer l and the feature map from the previous layer is given by Eq. (12), where p_l and s_l indicate the padding and stride values for the l-th layer, respectively. If the size is not divisible evenly, the convolutional layer truncates the size, while the pooling layer performs rounding up.

$$\begin{aligned} h_l &= \frac{h_{l-1} - k_l + 2p_l}{s_l} + 1, \\ w_l &= \frac{w_{l-1} - k_l + 2p_l}{s_l} + 1 \end{aligned} \tag{12}$$

3 Experiment Results

This section outlines the experimental setup, including the key parameters, simulation environment, evaluation criteria, and datasets used. We then present the results of the proposed approach, followed by an in-depth analysis of the findings.

3.1 Experimental Parameters

We set the learning rate to 1×10^{-3} to strike a balance between fast convergence and training stability. A higher learning rate may cause instability during training, whereas a lower value could result in unnecessarily prolonged training. To mitigate overfitting, we employ a weight decay coefficient of 5×10^{-4} for regularization, penalizing large weights and promoting simpler models that generalize better to unseen data. The model is trained for 50 epochs, a number selected based on monitoring both training and validation losses to ensure sufficient learning without overfitting. A batch size of 32 is used during training, balancing efficiency and memory constraints; smaller batches introduce noise in gradient updates, while larger ones may exceed memory capacity.

The integration of these hyperparameters is crucial to the training process and significantly impacts the model's performance in the authentication task. This section first describes the source of the channel dataset, followed by a comparison of the performance of SE-ConvNet and a traditional CNN. During testing, both the SE-ConvNet and CNN models are evaluated by feeding them the test set. The comparison not only involves the probability of the received channel data belonging to the legitimate device (Alice), but also includes a series of other performance metrics.

3.2 Performance Metrics

Recall and False Positive Rate (FPR) are critical metrics for assessing the effectiveness of PLA schemes. Recall, also referred to as True Positive Rate (TPR), indicates the percentage of actual positive instances correctly identified by the model. Conversely, FPR reflects the model's propensity to misclassify negative samples. A low FPR indicates better discrimination between positive and negative samples. This is particularly important in applications such as security monitoring and fraud detection, where minimizing false alarms is critical. Precision, on the other hand, measures the proportion of samples labeled as positive by the model that are truly positive. The formulas for these metrics are given below:

$$\begin{aligned}
&\text{Recall(TPR)} = \frac{TP}{TP + FN},\\
&\text{FPR} = \frac{FP}{FP + TN},\\
&\text{Precision} = \frac{TP}{TP + FP}
\end{aligned} \tag{13}$$

In these equations, *FP* denotes instances where Bob incorrectly classifies Eve's channel matrix as authenticated. *TN* indicates the number of illegitimate labels that Eve's signal should have received. *FN* represents the number of legitimate channels that Bob misclassifies as illegitimate, while *TP* signifies the correct identification of legitimate tags from legitimate devices. The accuracy of the model is computed using the following formula:

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

To comprehensively assess the overall performance of SE-ConvNet, this study also introduces the F1-score, which combines accuracy and recall. The F1-score offers a comprehensive assessment of the model's performance, with values ranging from 0 to 1, where a higher score signifies better performance. The formula for the F1-score is as follows:

$$\text{F1-score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{15}$$

The loss function used for training the model is the binary cross-entropy loss, expressed as:

$$L_{\text{cross}} = -\left[\mathscr{C} \log \hat{\mathscr{C}} + (1 - \mathscr{C}) \log(1 - \hat{\mathscr{C}})\right] \tag{16}$$

In this equation, the classification labels for legitimate and illegitimate devices are assigned values of 1 and 0, respectively. The variable $\mathscr{C}$ represents the true label of the data, while $\hat{\mathscr{C}}$ indicates the predicted probability that the data comes from a legitimate device.

3.3 Dataset

In this study, we employed the advanced wireless channel generator QuaDRiGa [32] to simulate channel environments and obtain CSI data. The QuaDRiGa platform incorporates a drift model that enables the smooth evolution of small-scale parameters, such as multipath power, delay, departure angles, and arrival angles, over short time intervals as mobile terminals move along a specified trajectory.

The receiver's movement trajectory in the automobile assembly plant is illustrated in Fig. 6. Our experiment considers a 100 m × 100 m region, where a stationary receiver records channel data as the transmitter moves along different paths, introducing channel heterogeneity. CSI data gathered from various locations within this region are treated as originating from distinct emitters, thereby simulating the CSI of multiple transmitters.

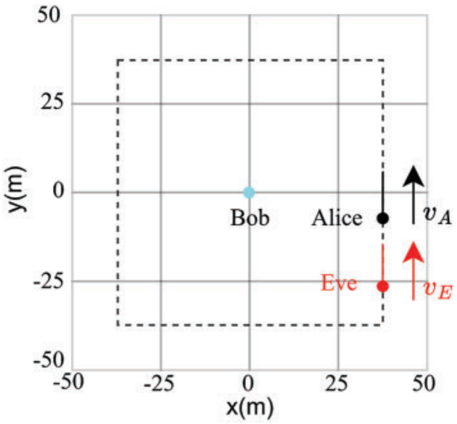

Figure 6: The receiver's position and the transmitter's movement trajectory in an automotive assembly plant

QuaDRiGa generated two datasets for this study: Dataset A, using parameters aligned with [17], and Dataset B, employing parameters tailored for experimental validation as described in Table 1. The QuaDRiGa platform generated a complex CSI matrix of size 10,000 × 1024, with each CSI instance represented as a 1 × 1024 complex vector. Each vector was subsequently converted into a 32 × 32 channel matrix. The dataset was partitioned into three subsets for training, validation, and testing, with a distribution of 80%, 10%, and 10%, respectively. For comparative evaluation, we selected a CNN architecture from [17] as a baseline model.

Table 1: Parameter settings for Dataset B

Options	**Parameter settings**
Channel model	3GPP_38.901_UMi_NLOS
Center frequency	3.5 GHz
Bandwidth	20 MHz
Base station height	10 m
User height	1.5 m
User movement speed	1 m/s
Number of subcarriers	32
Polarization direction	Vertical polarization
Antenna configuration	AP: 32 antenna Nodes: Single antenna

Since SE-ConvNet requires a specific data format for training and testing, the CSI data must be normalized. Normalization refers to mapping the input values to the range of $[0,1]$ to enhance the learning efficiency of the authenticator. The normalization formula is as follows:

$$\widehat{\mathbf{H}}^{l'}_{a_{\text{norm}}}(t) = \frac{\widehat{\mathbf{H}}^{l}_{a}(t) - \widehat{\mathbf{H}}_{a_{\min}}(t)}{\widehat{\mathbf{H}}_{a_{\max}}(t) - \widehat{\mathbf{H}}_{a_{\min}}(t)} \tag{17}$$

In the 1-D case, $\mathbf{H}_a(t)$ is transformed by obtaining its maximum $\widehat{\mathbf{H}}_{a_{\max}}(t)$ and minimum $\widehat{\mathbf{H}}_{a_{\min}}(t)$ values. The 1-D is denoted as $\widehat{\mathbf{H}}_a^{l}(t)$, and the normalized value is $\widehat{\mathbf{H}}_{a_{\text{norm}}}^{l'}(t)$. Moreover, before using the authenticator to identify devices in the wireless communication environment, the authenticator must be trained. Specifically, we input the preprocessed dataset into the authenticator for training to extract deep features, thereby acquiring an authentication model that can represent and differentiate spatiotemporal environmental features at different locations.

3.4 Comparative Experiments and Results Analysis

When tested on Dataset A, both CNN and the proposed SE-ConvNet achieved high accuracy, as shown in Table 2 below:

Table 2: Comparison of network models and authentication accuracy on Dataset A

Name	Network model	Accuracy (under 0 dB)
Our scheme	SE-ConvNet	96.9%
Scheme proposed by X. Li	CNN	94.8%

Fig. 7 illustrates the authentication accuracy comparison between SE-ConvNet and a traditional CNN across varying SNR conditions on Dataset B. When the SNR is 20 dB, both models achieve over 95% accuracy, demonstrating their effectiveness in high-quality signal environments. However, as the SNR drops to 0 dB, SE-ConvNet maintains high accuracy, whereas the CNN's performance declines significantly. The robustness of SE-ConvNet at 0 dB SNR can be attributed to the SE module, which enhances the model's ability to accurately process low-SNR channel data.

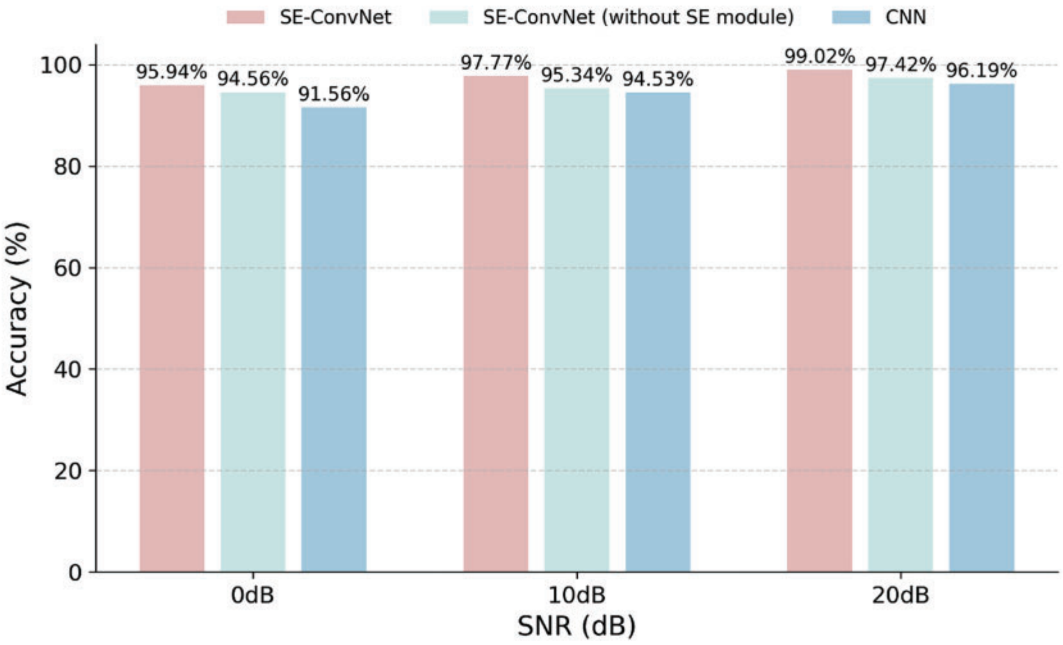

Figure 7: A comparison of authentication accuracy between PLA scheme based on SE-ConvNet and CNN

To further analyze the training process of SE-ConvNet, Fig. 8 illustrates the evolution of its accuracy and loss curves as the number of epochs increases under different SNR levels. The training and validation losses consistently decrease, while accuracy enhances, demonstrating that the model effectively fits the data with strong generalization ability. These results validate the training efficacy of SE-ConvNet under specific channel conditions and highlight its adaptability to the training data. Moreover, the stable trends suggest that the model avoids overfitting during training, providing a robust foundation for practical applications.

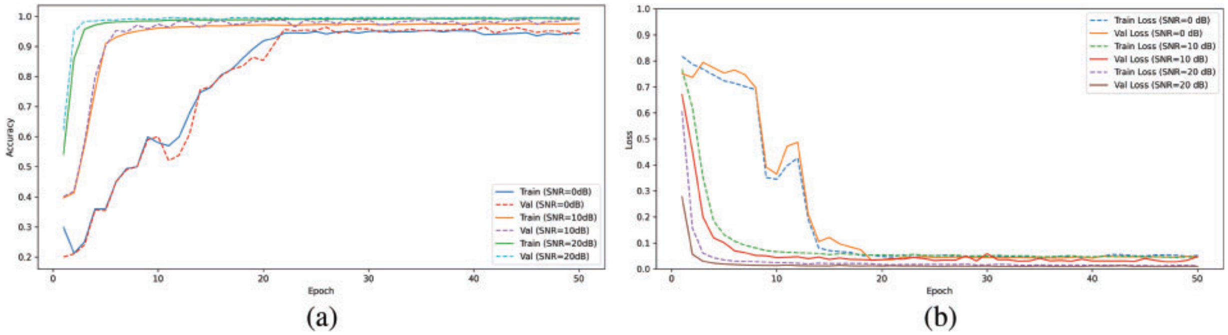

Figure 8: Accuracy and loss curves: (a) Accuracy curves; (b) Loss curves

Fig. 9a presents the ROC curve of SE-ConvNet under different signal-to-noise ratios (SNR). At an SNR of 0 dB, the area under the ROC curve (AUC) reaches 0.980, indicating SE-ConvNet's strong performance in distinguishing legitimate from illegitimate devices even in low-SNR environments. Fig. 9b illustrates the F1-score variations across different thresholds when the SNR is 0 dB. The F1-score remains above 0.9 for thresholds within the [0.5, 0.9] range, peaking near a threshold of 0.85, which signifies optimal classification accuracy under these challenging conditions.

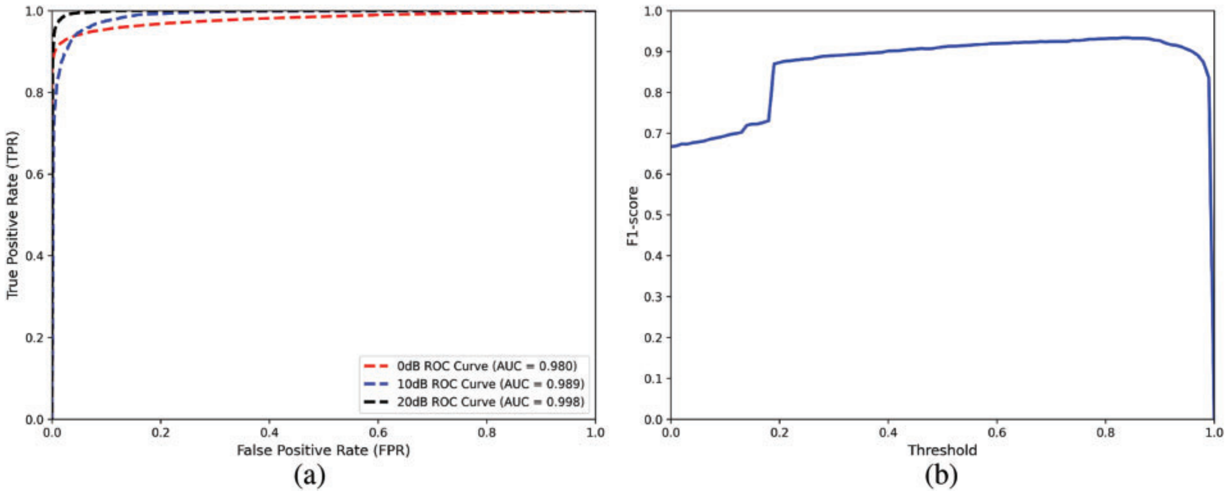

Figure 9: Classification performance metrics: (a) ROC curves; (b) F1-scores at 0 dB (SNR)

Fig. 10 provides a comparative analysis of FLOPS and memory usage between SE-ConvNet and CNN with an input data size of 2 × 32 × 32. FLOPS serves as a key indicator of a model's computational complexity, with lower values indicating greater efficiency in inference. Memory usage measures the amount of memory required for model parameters and intermediate activations during inference. Time per batch indicates the duration needed to process a specified number of input samples. A lower time per batch reflects higher throughput and faster response times, critical for real-time applications. Analyzing this metric helps identify optimization opportunities for improved computational efficiency. The comparison clearly shows that SE-ConvNet has significantly lower computational complexity than traditional CNN while maintaining accuracy. This finding not only demonstrates SE-ConvNet's advantages in resource-constrained environments but also indicates its feasibility for large-scale deployment.

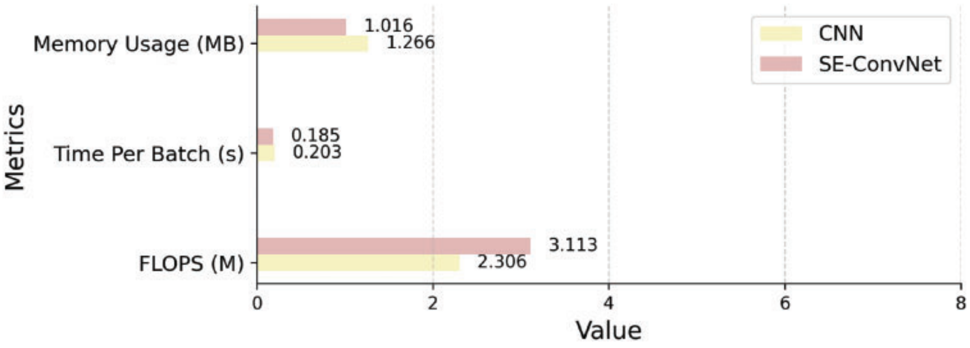

Figure 10: Model efficiency comparison

4 Conclusion

In this paper, we present SE-ConvNet, a lightweight scheme designed to detect physical layer spoofing attacks for mobile devices in IoT scenarios. This approach achieves packet authentication by leveraging the CSI of devices across various locations to accurately identify spoofing attacks. The simulation results show that, compared to other authentication schemes, the proposed SE-ConvNet exhibits superior robustness under time-varying conditions and environmental disturbances, while maintaining a lower model overhead. Notably, even at 0dB (SNR), the AUC remains above 0.980, highlighting the scheme's resilience under challenging signal conditions. The current approach may still experience performance degradation under certain class imbalance conditions in the dataset. Therefore, future work can address these issues through further improvements in data processing and model architecture.

Future research should focus on developing PLA schemes suitable for large-scale multi-user mobile scenario. As proposed in [27], optimizing feature extraction and selection of CSI samples enhances feature discriminability, improving PLA efficiency in complex communication environments. Additionally, exploring methods that adapt to multiple application scenario while balancing performance and complexity will be critical.

Acknowledgement: The authors would like to express their gratitude to the anonymous reviewers and the editorial team for their feedback and suggestions.

Funding Statement: This work was supported in part by the National Key R&D Program of China under grant no. 2022YFB2703000, in part by the Young Backbone Teachers Support Plan of BISTU under grant no. YBT202437, in part by the R&D Program of Beijing Municipal Education Commission under grant no. KM202211232012, and in part by the Educational Innovation Program of BISTU under grant no. 2025JGYB19.

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Xiaoying Qiu, Xiaoyu Ma; data collection: Xiaoyu Ma; analysis and interpretation of results: Xiaoying Qiu, Guangxu

Zhao; draft manuscript preparation: Xiaoying Qiu, Xiaoyu Ma, Guangxu Zhao, Jinwei Yu, Wenbao Jiang, Zhaozhong Guo, Maozhi Xu. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: All the experimental data are presented in this paper.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Jiang JR. Short survey on physical layer authentication by machine-learning for 5G-based Internet of Things. In: 2020 3rd IEEE International Conference on Knowledge Innovation and Invention (ICKII); 2020; Kaohsiung, Taiwan: IEEE. p. 41–4.
2. Illi E, Qaraqe M, Althunibat S, Alhasanat A, Alsafasfeh M, de Ree M, et al. Physical layer security for authentication, confidentiality, and malicious node detection: a paradigm shift in securing IoT networks. IEEE Commun Surv Tutor. 2024;26(1):347–88. doi:10.1109/COMST.2023.3327327.
3. Kaur B, Dadkhah S, Shoeleh F, Neto ECP, Xiong P, Iqbal S, et al. Internet of things (IoT) security dataset evolution: challenges and future directions. Internet Things. 2023;22(10):100780. doi:10.1016/j.iot.2023.100780.
4. Zhang F, Shi BX, Jiang WB. Review of key technology and its application of blockchain. Chin J Netw Inf Secur. 2018;4:22–9.
5. Fang H, Zhu Y, Zhang Y, Wang X. Decentralized edge collaboration for seamless handover authentication in zero-trust IoV. IEEE Trans Wirel Commun. 2024;23(8):8760–72. doi:10.1109/TWC.2024.3354064.
6. Dotan M, Pignolet YA, Schmid S, Tochner S, Zohar A. Survey on blockchain networking: context, state-of-the-art. chall ACM Comput Surv (CSUR). 2021;54(5):1–34.
7. Wang N, Wang P, Alipour-Fanid A, Jiao L, Zeng K. Physical layer security of 5G wireless networks for IoT: challenges and opportunities. IEEE Internet Things J. 2019;PP(99):1. doi:10.1109/JIOT.2019.2927379.
8. Wang N, Li W, Wang P, Alipour-Fanid A, Zeng K. Physical layer authentication for 5G communications: opportunities and road ahead. IEEE Netw. 2020;34(6):198–204. doi:10.1109/MNET.011.2000122.
9. Liu J, Wang X. Physical layer authentication enhancement using two-dimensional channel quantization. IEEE Trans Wirel Commun. 2016;15(6):4171–82. doi:10.1109/TWC.2016.2535442.
10. Xie N, Li Z, Tan H. A survey of physical layer authentication in wireless communications. IEEE Commun Surv Tutor. 2020;23(1):282–310. doi:10.1109/COMST.2020.3042188.
11. Wang HM, Fu QY. Channel-prediction-based one-class mobile IoT device authentication. IEEE Internet Things J. 2021;9(10):7731–45. doi:10.1109/JIOT.2021.3114348.
12. Xu Q, Zheng R, Saad W, Han Z. Device fingerprinting in wireless networks: challenges and opportunities. IEEE Commun Surv Tutor. 2016;18(1):94–104. doi:10.1109/COMST.2015.2476338.
13. Guo Y, Zhang J, Hong YWP. Deep learning-enhanced physical layer authentication for mobile devices. In: GLOBECOM 2023—2023 IEEE Global Communications Conference; 2023; Kuala Lumpur, Malaysia: IEEE. p. 826–31.
14. Hoang TM, Vahid A, Tuan HD, Hanzo L. Physical layer authentication and security design in the machine learning era. IEEE Commun Surv Tutor. 2024;26(3):1830–60. doi:10.1109/COMST.2024.3363639.
15. Liu R, Li Y, Zhang M, Ding Z, Yang S, Zhu S. The wireless IoT device identification based on channel state information fingerprinting. In: 2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC); 2020; Chongqing, China: IEEE. Vol. 9, p. 534–41. doi:10.1109/ITAIC49862.2020.
16. Liao R, Wen H, Pan F, Song H, Xu A, Jiang Y. A novel physical layer authentication method with convolutional neural network. In: 2019 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA); 2019; Dalian, China: IEEE. p. 231–5.
17. Li X, Huang K, Wang S, Xu X. A physical layer authentication mechanism for IoT devices. China Commun. 2021;19(5):129–40. doi:10.23919/JCC.2021.00.014.

18. Qiu X, Dai J, Hayes M. A learning approach for physical layer authentication using adaptive neural network. IEEE Access. 2020;8:26139–49. doi:10.1109/ACCESS.2020.2971260.
19. Wang S, Huang K, Xu X, Zhong Z, Zhou Y. CSI-based physical layer authentication via deep learning. IEEE Wirel Commun Lett. 2022;11(8):1748–52. doi:10.1109/LWC.2022.3180901.
20. Gao N, Huang Q, Li C, Jin S, Matthaiou M. Environment semantics enabled physical layer authentication. IEEE Wirel Commun Lett. 2024;13(1):178–82. doi:10.1109/LWC.2023.3324981.
21. Xie W, Wang H, Feng Z, Ma C. A novel PHY-layer spoofing attack detection scheme based on WGAN-encoder model. IEEE Trans Inf Forensics Secur. 2024;19(9):8616–29. doi:10.1109/TIFS.2024.3460373.
22. Chen Y, He H, Liu S, Zhang Y, Li Y, Xing B, et al. Physical layer authentication for industrial control based on convolutional denoising autoencoder. IEEE Internet Things J. 2024;11(9):15633–41. doi:10.1109/JIOT.2023.3347603.
23. Chen S, Pang Z, Wen H, Yu K, Zhang T, Lu Y. Automated labeling and learning for physical layer authentication against clone node and sybil attacks in industrial wireless edge networks. IEEE Trans Ind Inform. 2020;17(3):2041–51. doi:10.1109/TII.2020.2963962.
24. Martins J, Gomes M, Silva V, Dinis R. Deep learning-based channel prediction for wireless physical layer security. In: 2024 IEEE International Mediterranean Conference on Communications and Networking (MeditCom); 2024; Madrid, Spain: IEEE; p. 114–8.
25. Qiu X, Sun X, Hayes M. Enhanced security authentication based on convolutional-LSTM networks. Sensors. 2021;21(16):5379. doi:10.3390/s21165379.
26. Jing T, Huang H, Wu Y, Gao Q, Huo Y, Sun J. Threshold-free multi-attributes physical layer authentication based on expectation–conditional maximization channel estimation in Internet of Things. Int J Distrib Sens Netw. 2022;18(7):15501329221107822. doi:10.1177/15501329221107822.
27. Wang Q, Pang Z, Liang W, Zhang J, Wang K, Yang Y. Spatiotemporal gradient-based physical-layer authentication enhanced by CSI-to-image transformation for industrial mobile devices. IEEE Trans Ind Inform. 2024;20(3):4236–45. doi:10.1109/TII.2023.3316178.
28. Pan F, Li X, Pu H, Guo Y, Liu J. Physical layer authentication based on residual network for industrial wireless CPSs; 2020; Singapore: IEEE. p. 4368–73. doi:10.1109/IECON43393.2020.
29. Jing T, Huang H, Gao Q, Wu Y, Huo Y, Wang Y. Multi-user physical layer authentication based on CSI using ResNet in mobile IIoT. IEEE Trans Inf Forensics Secur. 2024;19:1896–907. doi:10.1109/TIFS.2023.3340090.
30. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2016; Las Vegas, NV, USA: IEEE. p. 770–8.
31. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2018; Salt Lake City, UT, USA: IEEE. p. 7132–41.
32. Jaeckel S, Raschkowski L, Börner K, Thiele L. QuaDRiGa: a 3-D multi-cell channel model with time evolution for enabling virtual field trials. IEEE Trans Antennas Propag. 2014;62(6):3242–56. doi:10.1109/TAP.2014.2310220.