EURASIP Journal on Advances in Signal Processing

RESEARCH Open Access

Active eavesdropping detection: a novel physical layer security in wireless IoT

Mingfang Li[1*] and Zheng Dou[1]

*Correspondence: limingfang@hrbeu.edu.cn

[1] College of Information and Communication Engineering, Harbin Engineering University, Nantong Street, Harbin 150001, China

Abstract

Considering the variety of Internet of Things (IoT) device types and access methods, it remains necessary to address the security challenges we currently encounter. Physical layer security (PLS) can offer streamlined security solutions for the next generation of IoT networks. Presently, we are witnessing the application of intelligent technologies including machine learning (ML) and artificial intelligence (AI) for precise prevention or detection of security breaches. Active eavesdropping detection is a physical layer security-based method that can differentiate wireless signals between wireless devices through feature classification. However, the operation of numerous IoT devices operate in environments characterized by low signal-to-noise ratios (SNR), and active eavesdropping attack detection during communication is rarely studied. We assume that the wireless system comprising an access point (AP), *K* authorized users and a proactive eavesdropper (E), following the framework of transforming wireless signals at AP into organized datasets that this article proposes a BP neural network model based on deep learning as a classifier to distinguish eavesdropping and non-eavesdropping attack signals. By conducting experiments under SNRs, the numerical results show that the proposed model has stronger robustness and detection accuracy can significantly improve the up to 19.58% compared with the reference approach, which show the superiority of our proposed method.

Keywords: Deep learning, Internet of things (IoT), Physical layer security, Active eavesdropping detection, BP neural network

1 Introduction

In recent years, global IoT devices have continued to be deployed on a large scale, with a strong growth in the number of connections, widely used in various fields of production and life. It is expected that by the end of 2023, more than 43 billion devices worldwide will be connected to the IoT. The cognitive allocation of spectrum resources and spectrum prediction in the IoT can ensure efficient communication between different devices and users, minimize interference and conflicts, and spectrum prediction is crucial for supporting the transmission of wireless communication signals [1, 2]. While engaging with a variety of different devices, the wireless channels, due to their open nature, can potentially be exploited by unauthorized devices within signal range to intercept and pilfer signals, thereby compromising the security of the wireless IoT system.

The IoT network has become the target of illegal attacks, which poses great risks and challenges to the security of wireless IoT systems. In order to ensure network security, intrusion detection for wireless access is necessary.

Wireless communication networks as a part of the IoT plays an important role in the IoT. However, with the development of wireless communication technology, any two users in a wireless network can freely establish a connection, which causes an increasing scarcity of spectrum resources but also leads to a wide range of information leakage and adversarial attack problems, such as eavesdropping attacks and spoofing attacks. It reduces network availability and information transmission security[3]. [4] proposed an information security transmission scheme for IoT terminals under the joint optimization of done trajectory and resource allocation in the presence of eavesdroppers. Traditional upper layer encryption technology has many limitations, so PLS has become a potential solution to this problem. PLS plays a critical role in wireless communication and distinguishes itself from cryptographic techniques. It serves as a complementary approach to upper-layer cryptographic methods. The fundamental concept of PLS revolves around harnessing the inherent properties and vulnerabilities of wireless channels to establish secure communication at the physical layer (PHY). Many research solutions have been proposed regarding PLS, [5] proposed cognitive user imitation attacks that occur during spectrum switching in cognitive radio networks and utilize artificial intelligence technology to make effective autonomous decisions. The PHY transmission of cognitive radio networks based on nonorthogonal multiple access faces dual threats of primary user interference and third-party eavesdropping. A unmanned aerial vehicle assisted covert communication model to address the security threats faced in the process of air to ground communication, and maximized average concealment rate under illegal interception and complete secure transmission was proposed in [6]. A new method in the study of physical layer security was discussed, which proposed probabilistic security features and the use of wireless maps to capture uncertainty in wireless environments, especially in eavesdropper channels [7]. In the presence of multiple active eavesdroppers, a closed form expression for the optimal power allocation strategy for transmitting signals and artificial noise (AN) was obtained and the minimum transmission power required to ensure reliable and secure communication was found in [8]. A new beamforming scheme to improve the reliability or security of the physical layer of integrated 5 G satellite networks was pointed in [9]. In order to improve the survivability and flexible scheduling ability of the 6th generation IoT drone wireless power supply communication system, a multi-drone trajectory and wireless scheduling resource scheme was designed in [10]. Most of the existing work emphasizes secure transmission strategies, evaluating the security performance of wireless transmission through secure interruption probability, positive confidentiality probability, and interception probability. [11] focused on solving security issues in IoT access authentication and proposed an optimized radio frequency fingerprinting (RFF) classification algorithm. [12] utilized RF characteristics to achieve efficient signal source recognition in limited resource situations, which is usually different from transmitting signals. [13] proposed physical layer security authentication method for wireless RF fingerprint recognition. [14] studied a new long tail specific radiation source identification method is proposed for

aircraft recognition. Most of these tasks are based on RF fingerprints to identify and authenticate physical layer devices. This is different from the active eavesdropping detection proposed in this article. Our goal is to detect the presence of eavesdroppers during wireless access, without focusing on the specific identity of the target. Therefore, eavesdropping detection for wireless communication signals is also crucial for physical layer security.

Security in the physical layer is generally defined through the concept of information theory. A common channel model for secure communication is the eavesdropping channel [15], where the sending node attempts to send confidential information and transmits it reliably to a legitimate receiving node while avoiding illegal attacks from eavesdropping nodes during transmission. In PLS, eavesdroppers usually are divided into two types of eavesdropping, namely passive eavesdropping that only listens without attacking and active eavesdropping that takes active attack who impersonates as legitimate users. Usually active eavesdroppers are more damaging because active eavesdropping initiates attacks that result in greater information leakage. The active eavesdropping attacks in conventional communication systems include frequency-conducting spoofing attacks and active eavesdropping attacks during data transmission. The former is different from the latter in that the eavesdropper (E) sends the same (frequency-conducting) signal that is synchronized with the legitimate user [16]. Active eavesdroppers during data transmission can enhance information eavesdropping by broadcasting their own continuous wave signals [17], and a scheme based on active eavesdropping with rotated jamming to achieve wireless surveillance is considered in [18], where a legitimate eavesdropper (The reason why legitimate eavesdropper is mentioned here is because in this literature, eavesdropping link is used as legitimate surveillance link) performs information interception while the auxiliary jamming node interferes with the suspect link to successfully achieve legitimate eavesdropping. In recent years, due to energy constraints, for example in UAV communications, active eavesdropping techniques have received a lot of attention. It assumes that the suspicious link can detect wireless eavesdropping, covert surveillance is achieved by active eavesdropping [19]. It follows that there is an urgent need to implement effective anomaly detection with the aim of enhancing the reliability and availability of communication systems and to minimize the probability of interception.

This paper presents a novel approach where we put forward to build a neural network model to learn and classify datasets under the framework of deep learning, and introduce the discriminative loss functions into the training of the model. Compared with using the ML algorithm, the suggested approach demonstrates higher accuracy levels in learning data features, and the classification effect is significantly improved. Experiments have shown that reliable eavesdropping detection capabilities can be achieved for different eavesdropping attack scenarios by our solution approach.

The subsequent sections of this paper are arranged as follows: Section 2 is a summary of related works. Section 3 provides the system model. We introduce wireless communication systems with eavesdroppers and create a framework for wireless signal datasets. Section 4 describes the detection method based on machine learning. Section 5 presents a deep learning-based BP neural network eavesdropping detection scheme. Section 6 mainly analyzes the detection performance evaluation of different algorithms.

2 Related works

Intrusion detection is not only the main security problem addressed by the network layer, but also a problem that must be solved by the physical layer. Common intrusion detection algorithms include ML such as Bayesian networks, clustering analysis, support vector machine (SVM) and DL for example recurrent neural networks (RNN) and CNN for classification and prediction. This section primarily discusses detection approaches that utilize ML and DL techniques.

Machine learning techniques have been widely applied in anomaly detection, recognition, and text classification, and have achieved impressive results. Reference [20] discussed improving intrusion detection performance based on ML classifiers through feature selection in cyber-physical systems. Reference [21] proposed a new dual ended machine learning model to improve the prediction accuracy and real-time performance of heterogeneous spectral states. In a cognitive eavesdropping environment, Reference [22] adopted distributed machine learning algorithms are used to optimize the allocation ratio of secondary device resources to ensure the quality of service for users with higher task priorities. The author of [23] using the relationship between transmitted and received signals considering the transmission process to build a dataset and using SVM algorithm to classify eavesdropping and legitimate signals, but the detection accuracy of binary classification is not very high. Reference [24] utilized a lightweight network composed of BP neural network, auto regressive integral moving average model, and SVM to achieve intrusion detection and recognition. Reference [25] studied UAV wireless relay systems in the presence of active eavesdroppers and used single-class SVM and K-means clustering analysis to build predictive models to detect eavesdropping attacks, the study no longer considered general wireless systems but focused on UAV-assisted wireless systems, however, UAVs have limited energy and cannot directly detect eavesdropping attacks. Reference [26] used machine learning to process the actual propagation process of wireless signals and used Gaussian mixture model for classification. Reference [27] relied on Gaussian mixture model to identify spoofing attacks. Reference [28] considered reinforcement learning to detect spoofing attacks and achieve PLS authentication. Reference [29] utilized reinforcement learning algorithms in machine learning for spectrum sensing to quickly detect the required idle channels. Most of these works are based on ML for classification or detection but basically not really considered the actual detection capabilities of wireless access points.

Deep learning plays an important role in classification and recognition. Reference [30] studied the impact of adversarial attacks on device recognition based on CNN. Reference [31] proposed a dual denoising autoencoder approach to enhance the security of cyber physical systems by preventing eavesdropped. Reference [32] investigated the use of a bait detection scheme to bait malicious nodes to send fake routing responses to detect malicious nodes, and then using cryptographic encryption techniques to encode to avoid malicious eavesdropping. A defense strategy for spectrum sensing data forgery and eavesdropping hybrid attacks in Nakagami-m fading channels in cognitive radio networks is suggested as a means to attain both energy efficiency and physical layer security in [33]. A complex CNN for identifying signal spectrum information for multi-signal frequency domain detection and recognition is constructed in [34]. Using deep learning and few sample learning methods to identify different emitters based on RF fingerprint

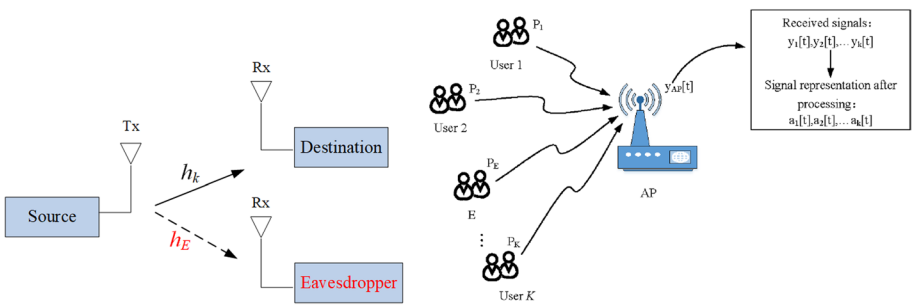
features in [35]. Classification tests are performed on public UCI datasets by deep convolutional neural networks and good performance is achieved in [36]. Implementation of wireless network security through self supervised learning and adversarial enhancement based few shot SEI method for transmitter authentication in [37]. Classification algorithms based on DL have progressively gained widespread acceptance as mainstream methodologies. However, deep learning is widely used in images, voice, video, text and other data classification problems such as radio signal recognition [38], where the data is public and more complex, and there is little research on classification of structured feature vector samples on eavesdropping attacks directly.

In this paper, we reframe the detection problem as a classification task to optimize the solution, which is traditionally solved based on ML algorithms. Therefore, this article mainly addresses the issue of active eavesdropping detection in the wireless access process of general wireless systems. In order to enhance the performance of eavesdropping detection as well as the accuracy of signal classification, we build on the idea of [23, 25] to generate test data from wireless signals, by using statistical knowledge of channel state information (CSI) to create a wireless signal dataset framework, and then artificial training data is created to input the data into ML and BP models. According to the characteristics of the dataset, a BP neural network model based on deep learning architecture has been proposed.

3 System model and creating dataset

3.1 System model

In this paper, the problem of active eavesdropping detection is studied. The classic eavesdropping model is shown in Fig. 1a. While the source node communicates with the destination node, there is a possibility that unauthorized eavesdroppers (referred to as E) may intercept the communication and employ deceptive techniques to mislead the destination node, thereby achieving their eavesdropping objectives. Our system model, as shown in Fig. 1b, a general wireless system that mainly solves the problem of active eavesdropping detection. The paper considers the system consisting of a single access point (AP), K authorized users, and active eavesdroppers (E) as seen in Fig. 1b. Each individual node is outfitted with a sole antenna, and the placement of them are randomized. The wireless channel connecting the AP and the k-th user is expressed as g_k. Likewise, the communication channel



(a) Wiretapping channel model. (b) System model.

Fig. 1 System model diagram

connecting the AP and E is expressed as g_E. Usually the wireless communication is divided into the uplink and downlink two phases.

In the uplink, the user sends the pilot sequence to the AP to request communication, and the AP performs channel estimation and identity authentication based on the pilot sequence. Assume that the pilot signal transmitted by the user k to the AP as $\mathbf{p}_k$, $\mathbf{p}_k \in \mathbb{C}^{\mathcal{L}\times 1}$ is referred to as a vector arranged in a column consisting of $\mathcal{L}$ entries, and $\|\mathbf{p}_k\|^2 = 1$. Among any two different users, that is when $k \neq k'$, $\mathbf{p}_k^{\dagger}\mathbf{p}_{k'} = 0$. If a malicious node E launches an attack to steal message s_k between user k and the AP, it will design the same pilot sequence $\mathbf{p}_E$ as the $\mathbf{p}_k$, and send it to the AP. At this time, the AP mistook it as the message request of two legitimate users. When the message is returned to the user, it will also be returned to E. Hence, the confidential information is inevitably leaked to E. That is, the SNR of the k-th user decreases as the power of E increases.

When E appears in the uplink and be proactive, it will result in a lower data rate for the user. Therefore, the disparity in data rates between user k and E, i.e., the channel capacity, becomes lower.

In the downlink transmission, the AP disseminates signals to the legitimates recipients. Of course, E will also receive these signals. It can be seen that the research on the detection problem of eavesdropping is very meaningful. If the existence of E is detected, we can stop the communication at any time or take confidential measures such as the convert transmission to reduce the risk of information leakage.

The focus of this study is the detectability of eavesdropping during the uplink communication, because accurate detection of eavesdropping can better realize attack identification and further complete identity authentication.

According to the spirit of the dataset framework cited in [23], the idea of using the correlation between the signal transmitted and the signal received to consider the transmission process is introduced into the representation learning of wireless signal features. When user k sends a message requesting communication to the AP, E will steal its message and imitate k while transmitting it to the AP. The only message available to the AP are received signals. At the t-th time slot, the signal received by the AP can be given by

$$y_{AP}[t] = \begin{cases} \sqrt{\mathcal{L}p_u}\sum_{k=1}^{K}\mathbf{p}_k g_k[t] + \mathbf{n}[t], & non-eavesdropping \\ \sqrt{\mathcal{L}p_u}\sum_{k=1}^{K}\mathbf{p}_k g_k[t] + \sqrt{\mathcal{L}p_E}\mathbf{p}_E g_E[t] + \mathbf{n}[t], & eavesdropping, \end{cases} \tag{1}$$

where $p_u \triangleq P_u/N_0, p_E \triangleq P_E/N_0$. In this equation, P_u and P_E refer to the mean transmitting power per user and E; N_0 represents the average noise power per receiving antenna; $\mathbf{n}$ is the additive white Gaussian noise (AWGN) vector with $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{\mathcal{L}})$. $y_{AP}[t]$, $g_k[t]$, $g_E[t]$ and $\mathbf{n}[t]$ are the representations of y_{AP}, g_k, g_E and $\mathbf{n}$ at time t.

3.2 Creating feature dataset

This section mainly creates a feature dataset framework. We use (1) to obtain the signal $y_{AP}[t]$ received at AP. Assuming that the pilot vector $\mathbf{p}_k^{\dagger}$ transformation processing with $y_k[t] = \mathbf{p}_k^{\dagger} y_{AP}[t]$, we can obtain $y_k[t]$,

$$y_k[t] = \begin{cases} \sqrt{\mathcal{L}p_u}g_k[t] + \mathbf{p}_k^{\dagger}\mathbf{n}[t], & non-eavesdropping \\ \sqrt{\mathcal{L}p_u}g_k[t] + \sqrt{\mathcal{L}p_E}g_E[t] + \mathbf{p}_k^{\dagger}\mathbf{n}[t], & eavesdropping. \end{cases} \tag{2}$$

Let $a_k[t] \triangleq |y_k[t]|^2$, then two values can be calculated at AP:

$$M_k^{(1)} \triangleq E_t a_k[t], \tag{3}$$

$$M_k^{(2)} \triangleq \frac{E_t a_k[t] - E_k|\mathbf{p}_k^{\dagger}\mathbf{n}[t]|^2}{E_k|\mathbf{p}_k^{\dagger}\mathbf{n}[t]|^2}, \tag{4}$$

according to sufficient statistical knowledge, it should be noted that $E_t\{\bullet\} \equiv E_{\{g_k\}_{k=1}^K}$, g_E, $\mathbf{n}^{\{\cdot\}}$ are dependent on $\{g_k\}_{k=1}^K$, g_E, $\mathbf{n}$ at t.

In the actual communication process, a user will send the pilot to the AP more than once when accessing the wireless network. Suppose that user sends the pilot vector T times, then there will be different values at AP, $a_k[1]$..... $a_k[T]$. According to (3) and (4), a dataset consisting of the following two eigenvalues can be created at AP:

Attribute 1 (Mean):

$$A_k^{(1)}[T] \triangleq \frac{1}{T}\sum_{t=1}^{T} a_k[t] = \begin{cases} A_{k|H_0}^{(1)}[T], & non-eavesdropping \\ A_{k|H_1}^{(1)}[T], & eavesdropping, \end{cases} \tag{5}$$

where H_0 indicates the state of absence of an eavesdropper, and H_1 indicates the state that there is an eavesdropper.

Attribute 2 (Ratio):

$$A_k^{(2)}[T] \triangleq \frac{\sum_{t=1}^{T} a_k[t] - \sum_{t=1}^{T}|\mathbf{p}_k^{\dagger}\mathbf{n}[t]|^2}{\sum_{t=1}^{T}|\mathbf{p}_k^{\dagger}\mathbf{n}[t]|^2} = \begin{cases} A_{k|H_0}^{(2)}[T], & non-eavesdropping \\ A_{k|H_1}^{(2)}[T], & eavesdropping. \end{cases} \tag{6}$$

When the AP acquires a substantial size of samples (i.e., when T reaches a significant value), then

$$A_k^{(1)}[T] \approx M_k^{(1)}, \tag{7}$$

Table 1 Features data: T points are related to eavesdropping, T points are not related to eavesdropping

Mean	Ratio
$A_{k\|H_0}^{(1)}[T_1]$	$A_{k\|H_0}^{(2)}[T_1]$
...	...
$A_{k\|H_0}^{(1)}[T]$	$A_{k\|H_0}^{(2)}[T]$
$A_{k\|H_1}^{(1)}[T_1]$	$A_{k\|H_1}^{(2)}[T_1]$
...	...
$A_{k\|H_1}^{(1)}[T]$	$A_{k\|H_1}^{(2)}[T]$

$$A_k^{(2)}[T] \approx M_k^{(2)}. \tag{8}$$

According to (2), (3) and (4), we convert the received signal to get the features data as the following tabular fashion of Table 1:

The training dataset starts from the T_1-th time slot data points, and when $T_1 = 1$, explain that all gathered data points during the uplink slots been used. Otherwise, $T_1 > 1$, only $T - T_1$ data points during the T_1-th to the T-th time period be used. The location of the t-th data point in the two-dimensional space can be described as $\left(A_k^{(1)}[t], A_k^{(2)}[t]\right)$. k indicates that we are detecting whether user k is under an eavesdropping attack.

According to (5) and (6), we can get labeled artificial training data sets for SVM algorithm and BP neural network model, as the following form of Table 2:

4 Eavesdropping detection with machine learning

ML and DL are common outlier detection algorithms, so before introducing our scheme, we first introduce the classic k-means++ and SVM methods that currently used, the ATD is inputted in them and our proposed method, then compare and analyze the detection performance of these three methods in experiments.

4.1 K-means++ clustering

Clustering algorithm is an unsupervised machine learning algorithm. K-means++ is an enhanced variant of the k-means clustering algorithm that is specifically devised to initialize cluster centers more efficiently, thereby improving clustering quality and algorithm performance. The k-means++ algorithm uses a smarter initialization method to select initial cluster centers to better represent the entire dataset, thereby reducing the risk of the algorithm falling into a local optimal solution. Because it's a binary classification problem (eavesdropping and non-eavesdropping), we set k=2. In this paper, ATD will be used for clustering model.

Specifically, the initialization procedure of the k-means++ algorithm proceeds obey the following manner:

1. Choose a sample from the dataset at random to serve as the initial cluster center.

Table 2 ATD: Labeled T points are related to eavesdropping, T points are not related to eavesdropping

Mean	Ratio	Labels
$A_{k\|H_0}^{(1)}[T_1]$	$A_{k\|H_0}^{(2)}[T_1]$	(0)
...	...	...
$A_{k\|H_0}^{(1)}[T]$	$A_{k\|H_0}^{(2)}[T]$	(0)
$A_{k\|H_1}^{(1)}[T_1]$	$A_{k\|H_1}^{(2)}[T_1]$	(1)
...	...	...
$A_{k\|H_1}^{(1)}[T]$	$A_{k\|H_1}^{(2)}[T]$	(1)

2. Compute the minimum distance among every data point and the current cluster center (i.e., the proximity to the closest cluster center), and choose the data point with the maximum distance conducting the next cluster center.
3. Iterate step 2 until k cluster centers are selected. It should be noted that the k-means++ algorithm requires significant computational resources due to its high computational complexity, especially when dealing with large-scale data sets.

4.2 SVM classifier

SVM is among the frequently utilized methods in ML. In the field of ML, it is a binary classifier with strong learning and generalization capabilities [39]. The fundamental model aims to establish a linear classifier that maximizes the margin within the feature space. In actual samples, many labeled samples are linearly indivisible, then the kernel trick can better handle this issue. In this paper, mainly conducts experiments with RBF kernel. The SVM takes the optimal separation hyperplane as the decision plane as Fig. 2 shown. Therefore, finding the maximum margin is the main optimization problem.

The optimal hyperplane optimization problem can be formulated by

$$\begin{aligned} &\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|^2 = 1 \\ s.t \quad &y_i\left(\left\langle \mathbf{w}^T \cdot \mathbf{x}_i \right\rangle + b\right) \geq 1,\ i = 1, ..., N, \end{aligned} \tag{9}$$

where N represents the total count of trained samples, $\mathbf{x}_i$ signifies the i-th sample data vector of the input. y_i denotes the i-th sample label. $y_i = +1$, if $\mathbf{x}_i$ is labelled as 0; otherwise, $y_i = -1$ if $\mathbf{x}_i$ as -1. The hyperplane shown in Fig. 2 is located between the two dividing lines, satisfying the $\langle \mathbf{w} \cdot \mathbf{x} \rangle + b = 0$. $1/\|\mathbf{w}\|^2$ representing the Euclidean distance from one edge to the hyperplane, the goal of (9) is to maximize the width of the two edges $2/\|\mathbf{w}\|^2$ to correctly separate samples.

For data that cannot be linearly separated, we introduce relaxation variable ξ_i, then we control the size of relaxation variables to achieve the optimal classification of the dataset.

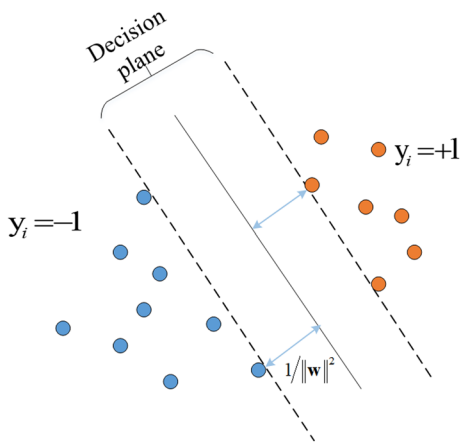


Fig. 2 Optimal hyperplane of SVM

The optimal hyperplane can be obtained through the optimization problem described as follows:

$$\begin{aligned} &\min_{\mathbf{w},b,\xi_n} \frac{1}{2}\|\mathbf{w}\|^2 = 1 + C\sum_{i=1}^{N}\xi_i \\ &s.t. \quad y_i\left(\mathbf{w}^T \cdot \mathbf{x}_i + b\right) \geq 1 - \xi_i,\ \xi_i \geq 0. \end{aligned} \tag{10}$$

C is the regularization coefficient. When $\xi_i = 0$, the point falls on the edge of the maximum interval, indicating that the data is correctly classified. If $0 < \xi_i < 1$, the points fall within the maximum interval, indicating that the data are correctly classified; as if $\xi_i = 1$, the point falls on the hyperplane and the data is correctly classified; otherwise data errors classification. According to the Lagrangian function and the KKT condition [23], the optimization problem can be described by (11)

$$\begin{aligned} &\max_{\lambda_1 \ldots \lambda_T} \sum_{i=1}^{N}\lambda_i - \frac{1}{2}\sum_{i=1,j=1}^{N}\lambda_i\lambda_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \\ &s.t. \quad \sum_{i=1}^{T} x_i y_i = 0 \quad 0 \leqslant \lambda_i \leqslant C,\ i \in \Omega\,. \end{aligned} \tag{11}$$

$K(\mathbf{x}_i, \mathbf{x}_j)$ denotes the vector inner product of the defined kernel function, mapping indistinguishable nonlinear data to a high-dimensional space for classification. λ_i is a Lagrange multiplier.

5 Eavesdropping detection with deep learning

In this section, we present a DL-based detection scheme using a BP neural network within the framework of deep learning. The paper first introduces deep learning into the wireless communication wiretapping detection.

Artificial neural network (ANN) can be described as a complex and interconnected system of adaptive neurons, the structure of ANN simulates mutually engage responses

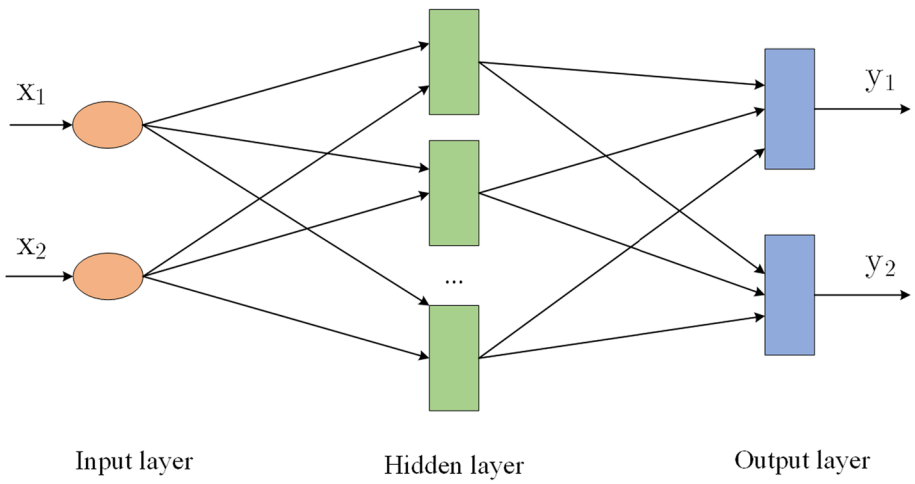


Fig. 3 BP neural network diagram

of the biological nervous system to visible stimuli. ANN possess the capacity for self-learning, self-organization, good error tolerant and outstanding nonlinear approximation capabilities. BP neural network is a type of multilayer feedforward neural network that training using the deviation backpropagation algorithm, and it is recognized as the most extensively employed neural network architectures.

As Fig. 3 shown, it is a typical 3-layer BP neural network diagram. No information interacted between neurons in the identical layer, and information is transferred between different layers according to the connection weight [40]. Its basic principle is to adjust the network weights and thresholds through the gradient descent algorithm, so that the discrepancy between the obtained output and the desired output is minimal or zero [41]. We designate the hidden layer to consist of 10 neurons, and use the output of this layer as the input of the next layer to attain the prediction outcome of the final classification, thus completing the construction of a typical BP neural network model. The BP neural network exhibits robust capabilities in nonlinear mapping and offers a versatile network structure. Using the BP model built above to train and classify the dataset, and adjust the constants associated with the feature values to achieve the best classification performance.

To enhance the accuracy of assessment while mitigating the risk of eavesdropping attacks, we conduct active wiretapping detection based on the model proposed above. As shown in Fig. 4 is the detection and training process based on wireless signal features, including the data processing phase and the model training and debugging phase. First, the wireless signal received during the wireless communication is processed to obtain a dataset of characteristic attributes and perform minimum-maximum normalization on it. Tensioning the dataset and input it into the BP model for training and testing. Then adjust the parameters of classifier selection and loss function and optimization function to fine-tune the model. Finally, the detection task of eavesdropping signal is completed.

The experiment is conducted using the PyTorch DL Framework as the implementation platform. 70% of the constructed dataset as the training examples for model training, and the leftover 30% as the testing data to test the performance of the model. During each model training process, different percentage divisions between the training and test sets can be set to enhance the authenticity and credibility of the results.

We select the sigmoid function as the Activation function. The Adam optimizer is employed for modifying the learning rate,and Softmax + Cross-entropy loss is used

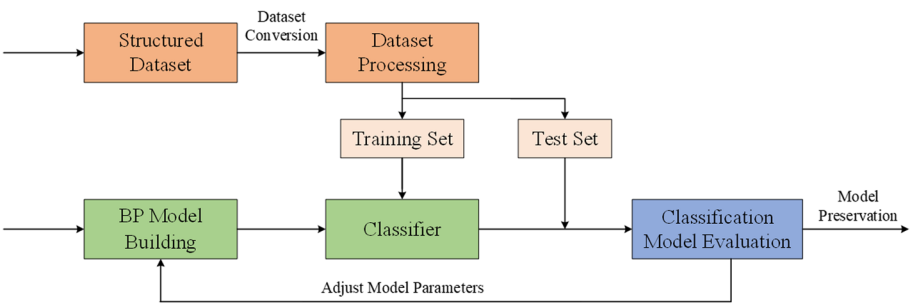


Fig. 4 BP neural network algorithm overall flow chart

as the classification loss function to expedite data calculation and make the numerical stability better. The softmax function is a discriminative function, and the output is a probability distribution. The difference in the probability of classification categories is more significant, and the form of the output distribution is closer to the real distribution. Here, the function expression is given as

$$Softmax(\mathbf{x}) = \frac{e^{x_i}}{\Sigma_j e^{x_j}}. \tag{12}$$

Cross-entropy loss is calculated as:

$$Loss = -\sum_{i=1}^{2} y_i \log \hat{y}_i, \tag{13}$$

where y is the true distribution, $\hat{y}_i$ is the network output distribution, j is the number of categories. Cross-entropy loss in the PyTorch deep learning framework combines softmax and cross-entropy to compute. The final calculation formula is as follows:

$$\begin{aligned} Loss(\mathbf{x}, class) &= -\log\left(\frac{exp(x[class])}{\Sigma_j exp(x[j])}\right) \\ &= -x[class] + \log\left(\Sigma_j exp(x[j])\right). \end{aligned} \tag{14}$$

So the last layer of model prediction, loss and learning processes can be expressed as Fig. 5:

6 Performance evaluations

In our experiment, we will give several numerical examples as specific condition settings. Considering that there may be some changes in the state of radio signals in the actual transmission environment [42], following the propagation of electromagnetic waves through various paths, Rayleigh fading is employed to represent the phenomenon of channel fading. Let $g_k = \sqrt{\beta_k} h_k, g_E = \sqrt{\beta_E} h_E, \beta_k$ and β_E represent the path loss. $h_k \sim \ell\mathcal{N}(\mathbf{0}, \mathbf{1}), h_E \sim \ell\mathcal{N}(\mathbf{0}, \mathbf{1})$. Set dataset parameters: $K = 4, \mathcal{L} = 10, p_u = p_E = 5, \beta_k = \beta_E = 1, T_1 = 1$. For the TC-SVM classifier, we choose the RBF kernel function, $K(x, x') = exp\left(\gamma \|x - x'\|_2^2\right)$, let $\gamma = 0.5$, C=1. Set DL learning rate: l=0.05.

In order to better measure and distinct performance in detecting anomalies of k-means++, SVM which are to compare and the BP neural network algorithm classification proposed in this paper, a comprehensive analysis has been conducted from the aspects of accuracy, precision, F1 score, ROC curve, Sun of Squared Errors (SSE), Silhouette Coefficient (SC) and Calinski-Harabasz Index (CK Index).

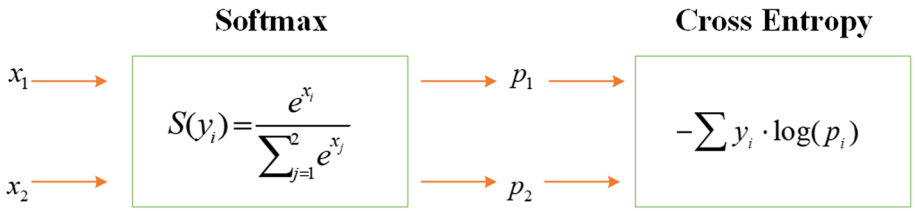


Fig. 5 BP neural network prediction, loss acquisition and learning process

Table 3 Comparison of the accuracy of different models

Model	*T*	n = 200 (%)	n = 2000 (%)	n = 4000 (%)
K-means++	10	62.50	61.67	63.33
	20	62.58	64.92	59.58
	50	57.83	43.88	58.63
SVM	10	76.25	82.50	83.75
	20	79.50	86.00	86.63
	50	80.87	91.56	89.00
BP	10	95.83	96.25	97.50
	20	94.98	96.20	98.84
	50	90.93	92.70	97.20

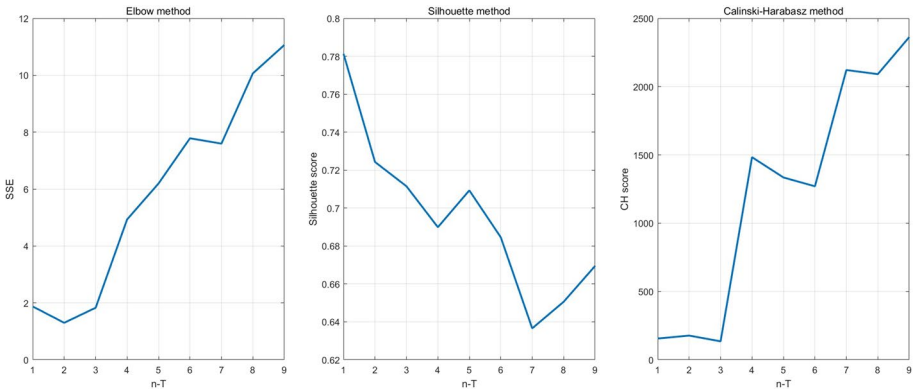


Fig. 6 SSE, SC and CH index in different cases

6.1 Comparison of accuracy of three algorithms under the same SNR

Accuracy serves as a metric to assess the performance of a classifier. We use the accuracy rate to judge the probability of accurately distinguishing eavesdropping signals from non-eavesdropping signals. It is calculated as

$$ACC = \frac{TP+TN}{TP+TN+FP+FN}, \tag{15}$$

where *TP* refers to instances that non-tapping samples are corrected classified, *TN* indicates that the eavesdropping samples are correctly classified, *FP* refers classifying non-tapping samples into eavesdropping samples and *FN* means classifying eavesdropping samples into non-tapping samples.

To bear out the effectiveness and accuracy of the BP model in wiretapping detection, this paper conducts comparative experiments on datasets in different scenarios on various models. The results of the experiments are presented in Table 3.

T represents the number of pilots sent by a single user, the size of samples (n) indicates the quantity of samples with a label of 0 (or label of 1). The feature composition of the dataset is composed of Mean and Ratio as shown in Table 2, and the experiment was conducted at SNR=10 dB. Comparing the results in Table 3, it becomes obvious that using BP neural network as a classifier is able to obtain higher detection

accuracy. We found that the k-means++ clustering effect is not ideal. So we use the common evaluation indexes of clustering algorithm SSE, SC and CK Index to measure the clustering effect of k-means++ algorithm. The results of the above nine data sets are shown in Fig. 6 (The x-axis $n - T$ refers to (200-10, 200-20, 200-50, 2000-10,... 4000-50)).

SSE is one of the most common metrics used to evaluate the effect of clustering, and its smaller value indicates better clustering effect. The SC is used to evaluate the tightness and separation of the clusters, and a larger value indicates that the cluster in which the data point is located is more reasonable. CH Index is a kind of index to evaluate the effect of clustering, and the higher value indicates the better clustering effect. According to Fig. 6, we can see that under different T conditions when n is the same, the SSE curve shows an overall upward trend, the SC curve shows an overall downward trend, and the CH curve shows an overall downward trend; When T is the same and under different n conditions, as T increases, the SSE value increases, the SC value decreases, and the CH value increases. Therefore, based on comprehensive analysis, the k-means++algorithm is not suitable.

Comparing SVM algorithm and BP neural network, we found that under the same sample size, the detection accuracy of both classifiers escalates with the raises of T, and the BP model exhibits superior detection performance compared to the SVM method. As the sample size increases, there is an overall inclination of the SVM detection accuracy is increasing, which has little effect on the detection results of BP model. We visualize the classification effect as shown in Fig. 7 where in the case of $n = 200$, $T = 10$. Fig (a) shows the classification results based on SVM algorithm for the reference we compared, while Fig (b) shows the classification performance of the proposed BP neural network. Based on the comparison, it is evident that the classification performance of the BP neural network classifier proposed in this article is better.

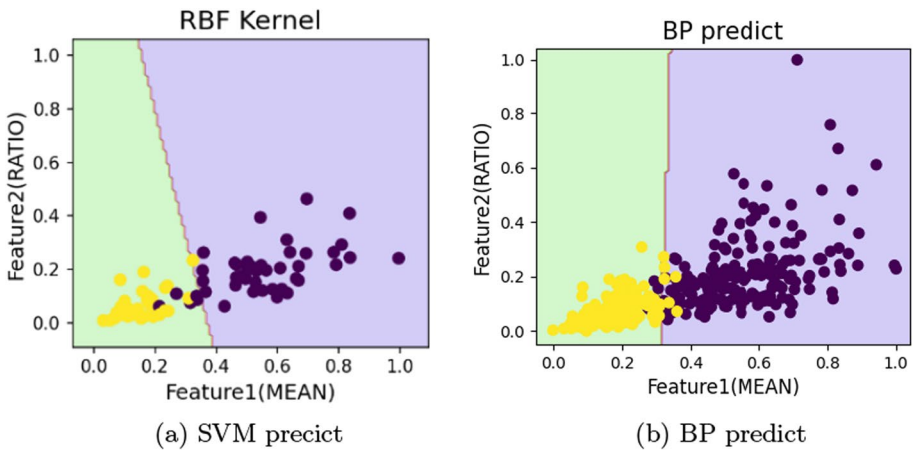
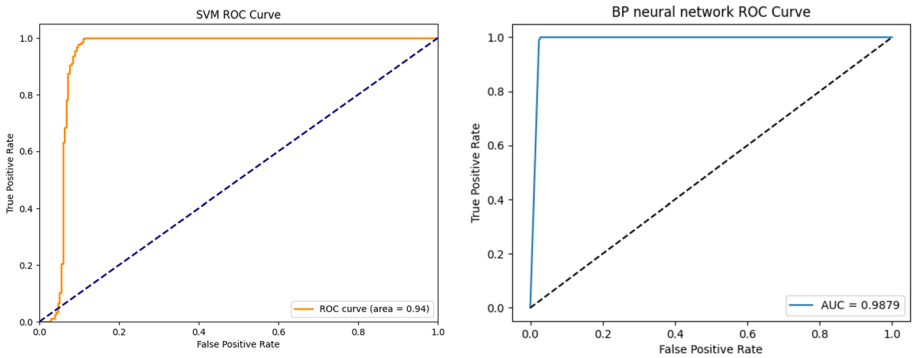


Fig. 7 SVM and BP classification comparison chart

Table 4 Comparison of experimental results with different indicators under different SNRs

SNR (dB)	Precision		F1 Score	
	SVM (%)	BP (%)	SVM (%)	BP (%)
−8	70.54	84.18	82.72	91.41
−6	70.54	80.15	82.72	88.98
−4	70.54	70.69	82.72	82.83
−2	70.54	71.07	82.72	83.41
0	71.79	69.02	83.58	81.67
2	70.92	72.14	82.98	83.82
4	73.01	71.19	84.40	83.17
6	73.71	70.36	87.45	82.60
8	74.67	93.11	85.50	96.43
10	78.52	94.30	87.97	97.07
12	84.40	95.92	91.54	97.92
14	83.33	96.36	90.90	97.83
16	83.98	98.85	91.29	99.42
18	88.67	98.23	93.99	93.11
20	86.26	97.50	92.62	98.73



(a) SVM ROC curve (b) BP neural network ROC Curve

Fig. 8 SVM and BP classification comparison chart

6.2 Comparison of precision and F1 scores between SVM and BP neural networks under different signal-to-noise ratios

Different from [22], in addition to comparing accuracy, we have also added new comparative indicators under different SNRs. When E launches an eavesdropping attack on the uplink, the SNR of the k-th user will decrease as the power of E increases. SNR can also be used to judge whether the signal is interfered. If the SNR is low, it means that the signal quality is poor and there is more noise in the signal. Therefore, SNR is a very important parameter that can serves as metrics for assessing algorithm performance. Hence, we analyze the behavior of the SVM algorithm and proposed in this paper by comparing the precision and F1 score under different SNR when n=2000, T=10. We summarize the experimental data in Table 4.

Precision measures the accuracy of the model in predicting non-tapping samples, the F1 score represents the balanced measure of precision and recall, calculated as

their harmonic mean, taking into account the accuracy and completeness of the model. According to the results in Table 4, we can find that our proposed BP neural network is almost always higher than SVM in precision and F1 score. To further enhance the validity of the model in this paper, we give the ROC curve plot for one of the cases as shown in Fig. 8.

On the ROC curve, the Area Under the Curve (AUC) indicates the performance of the algorithm, and the effectiveness is indicated by a higher AUC value, with a value closer to 1 indicating superior performance. ROC curve can help us make more objective decisions in model evaluation and selection.

It can be seen that the classification effect of the BP neural network classifier proposed in this paper is better. Comprehensive analysis shows that the deep learning-based BP neural network model performs satisfactorily, and it performs better in wireless communication eavesdropping detection scenarios.

7 Conclusion and future work

This paper mainly focuses on the research of physical layer security issues. Unlike existing work, we did not directly mention performance indicators for physical layer security, such as confidentiality capacity, interruption probability, and interception probability. The main goal of active eavesdropping attacks is to obtain sensitive information about communication by monitoring network traffic and intercepting wireless signals, without being observed by both parties, which undermines the confidentiality of communication. Therefore, the paper focuses on the research of active eavesdropping detection in the physical layer and models the detection problem as a signal classification problem, mainly focusing on the classification performance of the model for the presence or absence of eavesdroppers. By detecting active eavesdropping attacks, communication can be stopped in a timely manner to further adopt more effective transmission strategies, which can further prevent more serious attacks such as man in the middle attacks. Therefore, excellent active eavesdropping attack detection performance can help improve the confidentiality of physical layer secure communication.

In this paper, a new methodology to detect active wiretapping is presented for physical layer security based on deep learning. The method uses the BP neural network model to learn and classify the structured eigenvectors of wireless communication signals, and optimizes different parameters in the detection task to raise the active wiretapping detection property of the algorithm. Compared to the references, we have added experiments under different SNR conditions, and the model's effectiveness is validated through extensive experimentation on diverse datasets, yielding compelling results since the characteristics of the data itself will have a more obvious impact on the classification effect.

In the real world of the IoT wireless network, the access and exit of user devices are constantly changing, and the channel environment is also constantly changing. This article considers the situation based on static, which is also the limitation in the application of the model, which cannot be fully adaptive learning. In the wireless communication scenario proposed in this article, there is no limit on the number of access users. Therefore, in order to better adapt to future research on physical layer security issues in

wireless communication, we will collect real-world wireless signals to further verify the identity authentication problem of user devices in dynamic access and exit environments to improve the detection performance of active physical layer attacks. We will also continuously improve the model by introducing new mechanisms such as incremental learning and attention mechanisms to achieve adaptive dynamic learning in large-scale complex communication electromagnetic environments.

If any of the sections are not relevant to your manuscript, please include the heading and write 'Not applicable' for that section.

Abbreviations

IoT	Internet of things
PLS	Physical layer security
ML	Machine learning
AI	Artificial intelligence
DL	Deep learning
CNN	Convolutional neural network
SNR	Signal-to-noise
BP	Backpropagation
PHY	Physical layer
STBC	Space time block coding
AN	Artificial noise
RFF	Radio frequency finger
E	Eavesdropper
UAV	Unmanned aerial vehicle
SVM	Support vector machine
RNN	Recurrent neural network
ATD	Artificial training dataset
CSI	Channel state information
AP	Access point
ANN	Artificial neural network
SSE	Sum of squared errors
CK Index	Calinski-Harabasz Index
SC	Silhouette coefficient
AUC	Area under the curve

Acknowledgements
Not applicable.

Author contributions
ML proposed the original concept of the paper and completed the manuscript and revisions. ZD contributed to the review, editing and supervision.

Funding
None.

Availability of data and materials
The data supporting of the findings are included in the article. Please contact the author if any help needs.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Received: 16 September 2023 Accepted: 3 November 2023
Published online: 22 November 2023

References

1. X. Liu, H. Ding, S. Hu, Uplink resource allocation for NOMA-based hybrid spectrum access in 6G-enabled cognitive internet of things. IEEE Internet Things J. **8**(20), 15049–15058 (2021)
2. S. Li, Y. Sun, Y. Han, Y. Tu, O. Alfarraj, A. Tolba, P. Sharma, A novel joint time-frequency spectrum resources sustainable risk prediction algorithm based on TFBRL network for the electromagnetic environment. Sustainability **15**(6), 4777 (2023)
3. Y. Lin, H. Zhao, X. Ma, Y. Tu, M. Wang, Adversarial attacks in modulation recognition with convolutional neural networks. IEEE Trans. Reliab. **70**(1), 389–401 (2021)
4. Z. Na, C. Ji, B. Lin, N. Zhang, Joint optimization of trajectory and resource allocation in secure UAV relaying communications for internet of things. IEEE Internet Things J. **9**(17), 16284–16296 (2022)
5. G. Rathee, N. Jaglan, S. Garg, B.J. Choi, D.N.K. Jayakody, Handoff security using artificial neural networks in cognitive radio networks. IEEE Internet Things Mag. **3**(4), 20–28 (2020)
6. Y. Liu, Z. Na, Y. Zhang, X. Qin, B. Lin, Multi-UAV-assisted covert communications for secure content delivery in internet of things. Comput. Commun. **210**, 138–146 (2023)
7. Z. Utkovski, P. Agostini, M. Frey, I. Bjelakovic, S. Stanczak, Learning radio maps for physical-layer security in the radio access. In: 2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Cannes, France, pp. 1–5 (2019)
8. Y. Wu, R. Schober, D.W.K. Ng, C. Xiao, G. Caire, Secure massive MIMO transmission with an active eavesdropper. IEEE Trans. Inf. Theory **62**(7), 3880–3900 (2016)
9. M. Lin, Q. Huang, C. Tomaso et al., Integrated 5G-satellite networks: a perspective on physical layer reliability and security. IEEE Wirel. Commun. **27**(6), 152–159 (2020)
10. J. Wang, Z. Na, X. Liu, Collaborative design of multi-UAV trajectory and resource scheduling for 6G-enabled internet of things. IEEE Internet Things J. **8**(20), 15096–15106 (2021)
11. F. Xie, H. Wen, Y. Li et al., Optimized coherent integration-based radio frequency fingerprinting in internet of things. IEEE Internet Things J. **5**(5), 3967–3977 (2018)
12. Y. Lin, H. Zha, Y. Tu, S. Zhang, W. Yan, C. Xu, GLR-SEI: green and low resource specific emitter identification based on complex networks and fisher pruning. IEEE Trans. Emerg. Top. Comput. Intell. https://doi.org/10.1109/TETCI.2023.3303092.
13. D. Nouichi, M. Abdelsalam, Q. Nasir, S. Abbas, IoT devices security using RF fingerprinting. In: 2019 Advances in Science and Engineering Technology International Conferences (ASET), Dubai, United Arab Emirates, pp. 1–7 (2019)
14. H. Zha, H. Wang, Z. Feng et al., LT-SEI: Long-tailed specific emitter identification based on decoupled representation learning in low-resource scenarios. IEEE Trans. Intell. Transp. Syst. https://doi.org/10.1109/TITS.2023.3308716
15. A. Wyner, The wire-tap channel. Bell Syst. Tech. **54**(8), 1355–1387 (1975)
16. J.K. Tugnait, Pilot spoofing attack detection and countermeasure. IEEE Trans. Commun. **66**(5), 2093–2106 (2018)
17. B.-Q. Zhao, H.-M. Wang, P. Liu, Safeguarding RFID wireless communication against proactive eavesdropping. IEEE Internet Things J. **7**(12), 11587–11600 (2020)
18. H. Xu, L. Sun, Wireless surveillance via proactive eavesdropping and rotated jamming. IEEE Trans. Veh. Technol. **68**(11), 10713–10727 (2019)
19. S. Tu, J. Si, Z. Cheng, L. Guan, C. Wang, Performance of covert surveillance via proactive eavesdropping. In: 2020 International Conference on Wireless Communications and Signal Processing (WCSP), Nanjing, China, pp. 1–6 (2020)
20. S.E. Quincozes, D. Mosse, D. Passos, C. Albuquerque, L.S. Ochi, V.F. dos Santos, On the performance of GRASP-based feature selection for CPS intrusion detection. IEEE Trans. Netw. Serv. Manag. **19**(1), 614–626 (2022)
21. X. Liu, Q.Q. Sun, W.D. Lu, C.L. Wu, H. Ding, Big-data-based intelligent spectrum sensing for heterogenfous spectrum communications in 5G. IEEE Wirel. Commun. **27**(5), 67–73 (2020)
22. Y. Guo, R. Zhao, S. Lai, L. Fan, X. Lei, G.K. Karagiannidis, Distributed machine learning for multiuser mobile edge computing systems. IEEE J. Sel. Top. Signal Process. **16**(3), 460–473 (2022)
23. T.M. Hoang, T.Q. Duong, H.D. Tuan, S. Lambotharan, L. Hanzo, Physical layer security: detection of active eavesdropping attacks by support vector machines. IEEE Access **9**, 31595–31607 (2021)
24. Z. Zhao, Q. Du, H. Song, Traffic-load learning towards early detection of intrusion in industrial MMTC networks. IEEE Trans. Ind. Inf 1–10 (2022)
25. T.M. Hoang, N.M. Nguyen, T.Q. Duong, Detection of eavesdropping attack in UAV-aided wireless systems: unsupervised learning with one-class SVM and K-means clustering. IEEE Wirel. Commun. Lett. **9**(2), 139–142 (2020)
26. A. Weinand, M. Karrenbauer, R. Sattiraju, H. Schotten, Application of machine learning for channel based message authentication in mission critical machine type communication in processing. 23rd Eur. Wireless, Dresden, Germany, pp. 342–346 (2017)
27. X. Qiu, T. Jiang, S. Wu, M. Hayes, Physical layer authentication enhancement using a Gaussian mixture model. IEEE Access **6**, 53583–53592 (2018)
28. L. Xiao, Y. Li, G. Han, G. Liu, W. Zhuang, PHY-layer spoofing detection with reinforcement learning in wireless networks. IEEE Trans. Veh. Technol. **65**(12), 10037–10047 (2017)
29. X. Liu, C. Sun, M. Zhou, C. Wu, B. Peng, P. Li, Reinforcement learning-based multislot double-threshold spectrum sensing with bayesian fusion for industrial big spectrum data. IEEE Trans. Ind. Inf. **17**(5), 3391–3440 (2021)
30. Z. Bao, Y. Lin, S. Zhang, Z. Li, S. Mao, Threat of adversarial attacks on DL-based IoT device identification. IEEE Internet Things J. **9**(11), 9012–9024 (2021)
31. S. Wu, Y. Jiang, H. Luo, X. Li, Deep learning-based defense and detection scheme against eavesdropping and typical cyber-physical attacks. In: 2021 CAA Symposium on Fault Detection, Supervision and Safety for Technical Process (SAFERPROCESS), Chengdu, China, pp. 1–6 (2021)
32. M. Umar, A. Sabo, A.A. Tata, Modified cooperative bait detection scheme for detecting and preventing cooperative blackhole and eavesdropping attacks in MANET. In: 2018 International Conference on Networking and Network Applications (NANA), Xi'an, China, pp. 121–126 (2018)
33. X. Xu, Y. Wang, W.-P. Zhu, X. Hong, Y. Zou, Energy-efficient defensive strategy against hybrid SSDF/eavesdropping attacks over Nakagami-M Channels. IEEE Commun. Lett. **22**(4), 856–859 (2018)

34. C. Hou, G. Liu, Q. Tian, Z. Zhou, L. Hua, Y. Lin, Multisignal modulation classification using sliding window detection and complex convolutional network in frequency domain. IEEE Internet Things J. **9**(19), 19438–19449 (2022)
35. Z. Yao, X. Fu, L.T. Guo et al. Few-shot specific emitter identification using asymmetric masked auto-encoder. IEEE Commun. Lett. https://doi.org/10.1109/LCOMM.2023.3312669(2023)
36. Q.W. Dong, Application of Keras-based deep learning algorithm in structured data classification. J. JIAMUSI Univ. (Natural Science Edition). 40(04):47–49+54 (2022)
37. C. Liu, X. Fu, Y. Wang, L.T. Guo, Y.C. Liu, Y. Lin, et al. Overcoming data limitations: a few-shot specific emitter identification method using self-supervised learning and adversarial augmentation. https://doi.org/10.1109/TIFS.2023.3324394 (2023)
38. Y. Tu, Y. Lin, H. Zha, J. Zhang et al., Large-scale real-world radio signal recognition with deep learning. Chin. J. Aeronaut. **35**(9), 35–48 (2022)
39. J. Yang, H. Gao, Cultural emperor penguin optimizer and its application for face recognition. Math. Probl. Eng. 2020 (2020)
40. M. Yu, G. Dong, R. Xu, G. Yu, Precision time base source calibration prediction model based on BP neural network. China Test 1–7 (2023)
41. W. Hu, H. Jia, Automobile fault diagnosis method based on improved AFSA optimized BP neural network. Mech. Des. Manuf. Eng. **51**(11), 77–80 (2022)
42. Y. Zhao, L. Ge, H. Xie et al., ASTF: visual abstractions of time-varying patterns in radio signals. IEEE Trans. Vis. Comput. Graph. **29**(1), 214–224 (2023)

Publisher's Note