

Received 25 November 2022, accepted 4 December 2022, date of publication 12 December 2022, date of current version 10 January 2023.

Digital Object Identifier 10.1109/ACCESS.2022.3228641

RESEARCH ARTICLE

Performance Evaluation System Based on Multi-Indicators for Signal Recognition

QINGMING LI¹, SONGLIN YANG², AND HU CHEN³ 

¹VES Technologies Corporation Ltd., Beijing 100192, China

²College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China

³Wuhan Second Ship Design and Research Institute, Wuhan 430064, China


Corresponding author: Hu Chen (chenhu_14@163.com)

ABSTRACT Nowadays, the electromagnetic environment is becoming more and more complex, and it is increasingly difficult to accurately identify electromagnetic signals. At the same time, the quality of models and algorithms largely determines the final result of electromagnetic signal recognition, and a comprehensive and holistic evaluation system is needed to assess the quality of model or algorithm performance in the field of electromagnetic signal recognition. To address this phenomenon, this paper uses different machine learning models and AdaBoost algorithms to identify 100 classes of WiFi radiation source electromagnetic signals. In this paper, the classification performance, complexity and robustness of the models are evaluated comprehensively in three aspects respectively, and a multi-dimensional evaluation system for signal recognition is constructed, and the differences between different model algorithms in signal recognition effects are compared and analyzed, and the correlations between each evaluation index in different dimensions are explored. The experimental results show that different machine learning models and algorithms have different recognition effects on electromagnetic signals, among which the recognition accuracy of ResNet can reach more than 90%, but its computational complexity is high and easily affected by noise. DNN has poor recognition effect and the highest computational complexity, but it is not easily affected by noise. The AdaBoost algorithm does not necessarily improve the recognition classification accuracy of the underlying classifier. The evaluation system established in this paper is meaningful for assessing the performance of models and algorithms.

INDEX TERMS Evaluation systems, machine learning, ensemble learning, signal recognition.

I. INTRODUCTION

As the electromagnetic field continues to develop, the types and numbers of radiation sources in the electromagnetic environment are increasing. The electromagnetic environment is becoming more complex and the accurate identification of electromagnetic signals becomes difficult. At the same time, model evaluation is crucial to judge the performance of the model. The underlying performance, robustness, security, interpretability, and deployability of a model or algorithm all affect the quality of an AI application. A model or algorithm needs to be thoroughly considered before it is applied to an AI product.

The associate editor coordinating the review of this manuscript and approving it for publication was Yu Liu .

The traditional electromagnetic signal recognition techniques cannot make accurate predictive classification for some complex classification problems. Moreover, in some traditional machine learning algorithms, pre-processing of electromagnetic signals is required, and this process has significant limitations in extracting deep features of electromagnetic signals. In contrast, some deep learning models have also been applied in the field of electromagnetic signal recognition, and ensemble learning algorithms are a way to improve the performance of the models. In this paper, a comprehensive and integrated evaluation of the effectiveness of a class of models and algorithms in electromagnetic signal recognition is presented. Traditional model validation methods usually construct a validation set that is disjoint from the training set, and then evaluate the basic performance of the model based on the signal recognition accuracy of the model

on the validation set. It is difficult for us to judge the accuracy of a model or algorithm just from the accuracy rate. Nowadays, models are gradually moving towards lightweighting, and the complexity of the model becomes the focus of evaluation [1]. Meanwhile, the property of maintaining stability even in the case of abnormal inputs is called robustness. There are a large number of uncertainties in the electromagnetic space, and the electromagnetic signal is subject to various disturbances. The robustness of many models or algorithms has a great impact on the actual prediction results, so the robustness of models or algorithms needs to be evaluated and the prediction defects of models or algorithms under certain anomalous data and data distribution changes need to be found as early as possible. The approach proposed in this paper allows comprehensive testing and evaluation of the basic and business performance of the model. For the generalization ability of the model or algorithm in the prediction phase, the method tests the feasibility and compliance of the model or algorithm to the business objectives, and then proposes a targeted optimization scheme.

Electromagnetic signal recognition techniques can be divided into two main aspects, a feature engineering based approach and a deep learning approach. In [2], a radio frequency fingerprint feature extraction method based on the accumulated distance of I/Q data components is proposed. In [3], a signal feature extraction algorithm based on fractal complexity is proposed. In [4], complex-valued networks are applied for modulation identification. In [5], a radio individual identification method combining downscaling and machine learning is proposed. In [6], the accuracy of different deep learning models for individual recognition of electromagnetic signals is compared, including traditional CNN, ResNet and VGG networks. In [7], a zero-shot learning for signal recognition is proposed. In [8], a modulated signal classification method based on sliding window detection and complex convolutional networks is proposed. In [9], the evaluation analysis yields the advantages of plural networks in the signal classification task for radiation sources, such as higher recognition accuracy and faster learning rate. In [10], a new method for cognitive signal recognition based on hybrid information entropy and D-S evidence theory is proposed. In [11], an improved neural network pruning technique is applied to automatic modulation classification of edge devices. In [12], the semi-supervised learning approach is applied to modulated signal classification. In [13], a semi-supervised learning of digital signal modulation classification based on generative adversarial networks is proposed. In [14], the proposed AdaBoost algorithm periodically adjusts the weights of weak learners that are not correctly screened. In [15], the AdaBoost algorithm is used for radar signal modulation recognition. In [16], ensemble learning was used in signal modulation recognition.

In terms of model evaluation, most studies have focused on a single performance aspect of the model. In [17], a data distribution model was designed for an adversarial classification task and a general approach to design robust

classifiers was proposed and then evaluated. In [18], a hypothesis-countermeasure based security evaluation mechanism is proposed to improve the security of the model by considering the variation of the training and test data distribution, and thus evaluating the impact of the attack. In [19], the impact of attack effects on DNN-based device identification is investigated. In [20], a constellation graph signal dataset is proposed and the classification performance of the model is tested for different training set percentages under simulated channel fading. In [21], the modulation classification performance is evaluated for complex environments. In [22], the communication speech signals were evaluated. In [23], a spectrum focused frequency adversarial attack is proposed and the performance of this attack on modulation recognition is evaluated.

In this paper, a deep learning method is used to identify individual electromagnetic signal radiation sources and build an evaluation system with different dimensions for WiFi electromagnetic signal dataset. The evaluation system provides a more comprehensive evaluation of the model and evaluates the effect of the model on the individual identification of electromagnetic signal radiation sources. At the same time, this paper investigates the effect of integrated learning on the individual identification of electromagnetic signal radiation sources and evaluates the integrated learning algorithm. The integrated learning algorithm and the deep learning model are compared and analyzed. When facing the complex electromagnetic environment, the model or algorithm can be better selected and optimized according to the evaluation index.

II. EVALUATION SYSTEM

In the field of electromagnetic signal recognition, the establishment of the evaluation system needs to satisfy certain principles. First of all, the evaluation metrics should meet certain logical relationships between each other, and they should reflect not only the characteristics of the machine learning model, but also the internal connection between the metrics. Each part of the system consists of certain indicators, which are independent of each other and connected with each other. The construction of the evaluation system is hierarchical, with layers of depth, forming an indivisible evaluation system. In the face of the future complex electromagnetic environment, many key performance indicators are becoming more and more stringent, requiring a full range of models and algorithms to be evaluated [24].

The design of the evaluation system and the selection of evaluation indicators must be based on the principle of scientificity. The evaluation system can objectively and truly reflect the characteristics of the machine learning model in the signal field and the real relationship between the indicators objectively and comprehensively. Each evaluation index should be typical and representative, and reflect the comprehensive characteristics of a specific field as much as possible. Even when the number of indicators is reduced, it is easy to calculate the data and meet the reliable evaluation results. The setting of the evaluation index system, the weight distribution

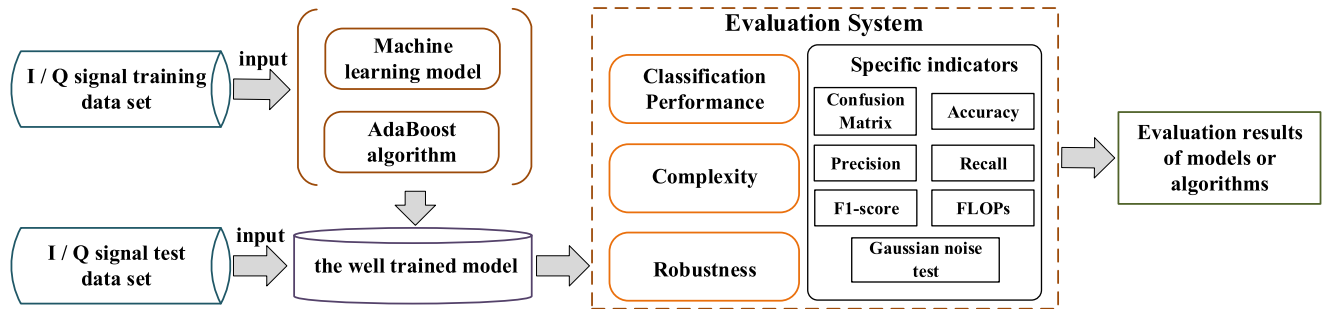


FIGURE 1. Evaluation system process.

of each indicator problem and the division of evaluation criteria should be combined with the conditions of the actual problem.

In the process of evaluating the electromagnetic signal recognition effect of machine learning models or algorithms, it is impossible to accurately and comprehensively evaluate machine learning models or algorithms by only evaluating the recognition accuracy of electromagnetic signals. Especially in the face of today’s complex electromagnetic environment, a multi-dimensional evaluation system is needed to evaluate the performance of machine learning models or algorithms. The innovation point of this paper is mainly in the comprehensive evaluation of evaluation indexes of different dimensions. Due to the characteristics of electromagnetic signals, suitable evaluation indexes need to be selected to evaluate the recognition performance of models and algorithms. In this paper, the evaluation system is divided into three dimensions of classification performance, complexity and robustness for comprehensive evaluation, and the recognition performance of the model and algorithm under different dimensions are evaluated respectively. Meanwhile, the relationship between the evaluation indexes of different dimensions is illustrated through the analysis of the experimental results.

The flow of the evaluation system in this paper is shown in Figure 1. Firstly, the machine learning model and AdaBoost algorithm are trained using the training set. Secondly, the I/Q signals from the test set are fed into the trained model or algorithm. Through the evaluation system, the performance of the models and algorithms are tested and evaluated comprehensively in terms of classification performance, complexity and robustness.

A. CLASSIFICATION PERFORMANCE

In signal recognition classification, confusion matrix, accuracy, precision, recall, F1 score, etc. are generally adopted as evaluation metrics for classification performance.

1) CONFUSION MATRIX

The confusion matrix provides an intuitive view of a model’s classification performance. As an evaluation method in supervised learning, it can visualize the predicted and true categories of a model in multi-categorization. In the matrix, each

Actual Class	Predicted Class	
	Positives	Negatives
Positives	TP (True Positives)	FN (False Negatives)
Negatives	FP (False Positives)	TN (True Negatives)

FIGURE 2. Confusion matrix.

row represents the predicted category of the model and each column represents the true labeled category.

In Figure 2, *TP* represents that the real result is a positive case and the predicted result is also a positive case; *FP* represents that the real result is a negative case and the predicted result is a positive case; *TN* represents that the real result is a negative case and the predicted result is also a negative case; *FN* represents that the actual result is a positive case and the predicted result is a negative case. The graph visually reflects the relationship between the predicted results of the model or algorithm and the true labels.

2) ACCURACY

Accuracy is one of the most commonly used metrics to assess the classification performance of a model and is usually used to describe how well a machine learning model classifies on all categories, often in cases where all categories are equally important. It describes the ratio between the number of all correct predictions and the total number of predictions, as shown by equation (1).

$$Accuracy = (TP + TN) / (TP + FN + FP + TN) \quad (1)$$

3) PRECISION

The precision is the ratio of the number of positive samples predicted correctly to the number of samples predicted to be positive, and the metric is primarily a measure of how accurately the model classifies samples as positive. It is shown as:

$$Precision = TP / (TP + FP) \quad (2)$$

4) RECALL

Recall mainly describes the ratio between the number of samples that are correctly predicted as positive samples and the total number of samples that are actually positive for the

disease. Recall and accuracy are intertwined and a balance between the two needs to be found based on actual needs. Recall is expressed as:

$$Recall = TP / (TP + FN) \quad (3)$$

5) F1-SCORE

F1-score is a comprehensive evaluation metric, which is a summation of precision and recall, and is closer to the smaller value between the two. It is more comprehensive to reflect the goodness of the system model. It is expressed as:

$$F1 = 2 (Precision \times Recall) / (Precision + Recall) \quad (4)$$

B. COMPLEXITY

Complexity is a common criterion for evaluating a model. The complexity of a model is related to the structural level of the model, internal parameters, and other factors. The higher the complexity of a model is not better. If a model has a high complexity, then the results in training may not be very good and the model appears to be underfitted. Therefore, for different practical problems and different training sets, the model needs to have a suitable complexity. In the case of similar classification and prediction performance of both models, the model with lower complexity is selected.

In this experiment, the complexity metric chosen is floating point operations (FLOPs), which is a proxy for the amount of computation and represents the number of floating point operations. FLOPs is commonly used to evaluate the complexity of a model (algorithm) and can reflect the computational power required by the model in the forward propagation process. This metric allows evaluating the hardware requirements of the model, including the required GPU performance and memory size.

During the training of a network model, many data processing operations are performed. There is a large amount of computational consumption in these operations. The overall computation of the model is equal to the sum of the computation of each operator in the model. Equations (5) and (6) reflect the way FLOPs are computed in the convolutional and fully connected layers of the network model.

At a layer of the convolutional layer with the number of input channels as C_{in} , the number of output channels as C_{out} , the convolutional kernel size as $K_w \times K_h$, and the size of the output feature map as $H \times W \times C_{out}$, the FLOPs is given by equation (5).

$$FLOPs = (M + A + 1) \times C_{out} \times H \times W \quad (5)$$

In the above equation, M is the number of multiplications, A is the number of additions. $+1$ is the bias, which reflects the error between the classification of the model on the sample data and the true label, and it reflects the accuracy of the model. $C_{out} \times H \times W$ is all the elements in the output feature map. The equations for M and A are as follows.

$$\begin{cases} M = C_{in} \times K_w \times K_h \\ A = C_{in} \times K_w \times K_h - 1 \end{cases} \quad (6)$$

In a fully connected layer, since there is no weight sharing, the number of FLOPs in that layer is the number of parameters in that layer. The specific expression for its calculation is shown by equation (7).

$$FLOPs = (2N_{in} - 1)N_{out} \quad (7)$$

In the above equation, N_{in} is the number of fully connected layer input nodes, N_{out} is the number of output nodes, $N_{in}N_{out}$ is the number of operations in which multiplication is performed, and $(N_{in} - 1)N_{out}$ is the number of operations in which addition is performed.

For the electromagnetic signal evaluation system proposed in this paper, the performance of machine learning models and AdaBoost algorithm for electromagnetic signal recognition needs to be evaluated. However, it can be seen from the FLOPs calculation formula that it is only applicable to the complexity assessment of machine learning models and not to some algorithms. Therefore, there is a certain threat to the validity of the evaluation system proposed in this paper.

C. ROBUSTNESS

The robustness of a model or algorithm is a test of how well it performs in the face of data anomalies. When data changes, how well the model or algorithm tolerates such data and whether the model's classification and prediction performance changes drastically. But robustness is not the same as stability, robustness is more about the ability to adapt to complex conditions. Electromagnetic signals may be affected by interference such as Gaussian noise, channel impairment, carrier frequency shift, etc., and therefore need to be tested for robustness [25].

SIGNAL-NOISE RATIO (SNR) represents the power ratio of signal to noise in a system. Its formula is shown by equation (8).

$$SNR = 10 \lg(P_s/P_n) \quad (8)$$

where P_s denotes the effective power of electromagnetic signal and P_n denotes the effective power of noise in dB.

III. CLASSIFICATION MODELS AND ALGORITHMS

A. DNN

The structure of Deep Neural Networks (DNN) is not fixed and mainly consists of an input layer, a hidden layer and an output layer [26].

Each layer has several neurons. Neurons between layers are interconnected, while neurons within layers are not interconnected, and neurons in the next layer connect all neurons in the previous layer. The model structure is shown in Figure 3. A neural network with more hidden layers is called a deep neural network.

Suppose the hidden layer is k , $k = 0, \dots, K-1$, the output vector of this layer is set to h^k , and the expression is shown as:

$$h_j^k = \sigma(x_j^k(v^k)) = \sigma((w_j^k)^T v^k + a_j^k) \quad (9)$$

In the above equation, the offsets and weight vectors of the nodes of neuron j in hidden layer k are denoted as a_j^k

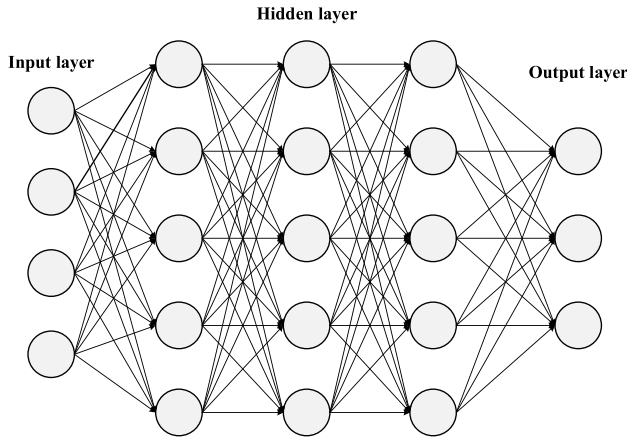


FIGURE 3. DNN model structure.

and $w_j^k \cdot \sigma(x_j^k)$ represents the sigmoid activation function with the expression $1/(1 + e^{-x_j^k})$. v^k represents the input vector of layer k . The output vector y^K of the last layer k of DNN is shown as:

$$y_j^K = \frac{e^{x_j^K(v^K)}}{\sum_j x_j^K(v^K)} = \text{soft max}_j(z^K(v^K)), \quad j = 1, \dots, N^K \quad (10)$$

B. CNN

Convolutional neural network (CNN) is a classical deep learning network model and an effective means to extract deep features using neural network theory. In the field of communication, CNN has also been applied to automatic signal recognition [27].

In the wireless domain, CNN does not operate on images, but on I/Q samples. In the I/Q plane, various radio signal waveforms exhibit different transition patterns, which can constitute the signal features that can eventually be learned by CNN filters. In the field of electromagnetic signal recognition, the network model requires a more comprehensive extraction of the fine and deep features of the electromagnetic signal, and generally saves the two I/Q signal data as two independent channels and splices them into a two-dimensional matrix, namely $Input = [N \times 2]$, where N represents the length of the electromagnetic signal. The electromagnetic signal data input vector is extremely “narrow” and generally satisfies $N \gg 2$. The structural parameters of CNN model is shown in Figure 4.

In the convolutional layer, the input data samples are processed to obtain the corresponding feature maps, and each convolutional kernel represents this kind of feature extractor. The features extracted by these feature extractors are not the same. A certain number of feature maps can be obtained through the continuous repetition of the convolutional layer. The output of the convolution layer is shown by equation (11).

$$Out_m = g(\sum W_m^n \otimes X^n + e_m) \quad (11)$$

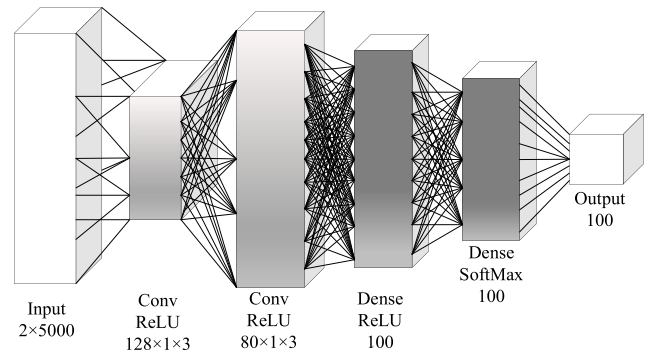


FIGURE 4. CNN model structure parameters.

In the above equation, Out_m is the output of feature map m . W_m^n is the convolutional kernel parameter, and the size of the parameter is set to control the convolutional kernel size and training efficiency running time, etc. \otimes is the convolutional operation. e_m is the bias of each convolutional layer. g is the activation function in the convolutional neural network.

C. AlexNet

The overall model contains five convolutional layers, two hidden fully connected layers and one output fully connected layer [28]. The model is able to extract deep features of data samples to a certain extent, and optimize the structural parameters of the model to reduce the computation of the model in the training process and improve the generalization ability. The Alexnet model parameters used in this paper are shown in Figure 5.

D. ResNet

For the traditional convolutional neural network, the model performance will gradually reach saturation and degradation problem as the number of layers of the network increases. The proposed residual module solves the network performance degradation problem to some extent by adding its residual module to this network[29]. It is shown in Figure 6. It differs from the traditional convolutional neural network structure in that it has a shortcut connection module.

Assuming that the fitted objective function is $H(x)$ and the nonlinear superposition layer is $F(x)$. The traditional approach is to make $F(x)$ close to $H(x)$, while the residual structure is to use $F(x)$ to approach $H(x)-x$. The advantage of this structure is that it allows the network to map shallow features to deep layers in a constant manner, and shallow and deep layers are effectively communicated. It can be seen that the output of the residual module is obtained by summing the output of several convolutional layers and the cascade between the input elements, and then the output of the residual unit is obtained by the ReLU activation function. The residual network contains many residual modules, which are cascaded together with each other.

The residual module structure as shown in Figure 6, which contains two convolutional layers and a shortcut connection

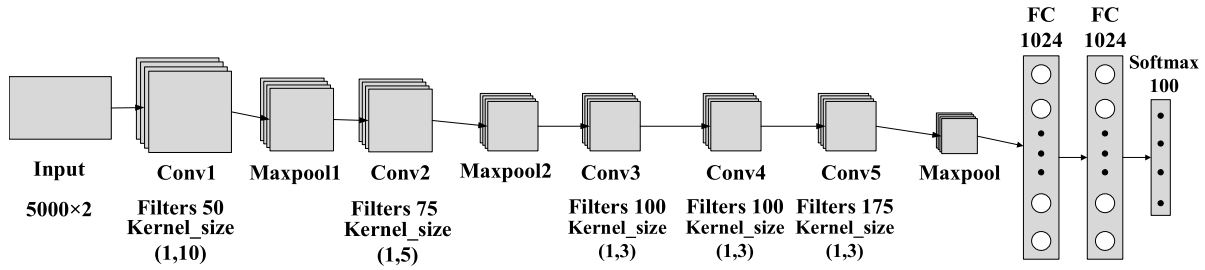


FIGURE 5. AlexNet model structure parameter.

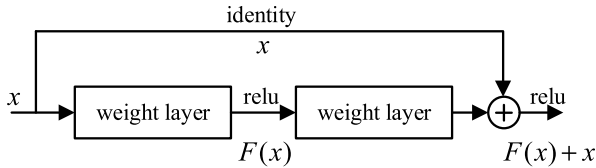


FIGURE 6. Residual unit.

method, the definition of the process is shown as:

$$y = F(x, \{W_i\}) + x \tag{12}$$

In the above equation, x and y are the input and output of the module, W_i is the weight of each layer, and $F(x, \{W_i\})$ is the mapping that the module needs to fit.

In the residual network, the ReLU function is chosen for the layer-to-layer activation function, and its expression is shown as:

$$f(x) = \max(0, x) \tag{13}$$

E. Adaboost

In the AdaBoost algorithm, the feature data training set is first given a total number of samples N as $\{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$, where each sample x_i represents a sample feature and each sample i corresponds to a sample label y_i . These sample labels together form a label set $Y, i = 1, 2, \dots, N$.

When a weak classifier is generated, it changes the weight D_{i+1} of the training set in the next weak classifier. When the conditions for the final output are satisfied, these weak classifiers are combined into a whole in a certain way to jointly determine the output[30].

First, the weights of the training samples are initialized, and the weights of each sample are taken as the mean value. It is expressed as:

$$D_1 = (w_{11}, w_{12}, \dots, w_{1N}) = (1/N, 1/N, \dots, 1/N) \tag{14}$$

Then iterate m times according to the artificially set values, $m = 1, 2, \dots, M$, where M is the total number of iterations. The base classifier h_m is trained based on the training set in the weight distribution D_m .

The error rate of h_m is calculated during the training process, and the formula is shown as:

$$e_m = \sum_{i=1}^N w_{mi} I(h_m(x_i) \neq y_i) \tag{15}$$

In the above equation, w_{mi} is the weight corresponding to feature data i at m iterations; I represents the classification result of weak classifier h_m on data x_i , if $h_m(x_i) \neq y_i$, then $I(h_m(x_i) \neq y_i) = 1$, otherwise $I(h_m(x_i) \neq y_i) = 0$.

At the end of training, the algorithm combines all base classifiers together by a linear weighting method. The base classifier coefficients a_m are selected, $a_m \in R$. Although a_m represents the weights of the base classifiers, its essence represents the importance of each base classifier and its specific expression is shown by equation (16).

$$a_m = \frac{1}{2} \ln \frac{1 - e_m}{e_m} \tag{16}$$

The weight distribution of the training set is continuously updated during the training learning process of the algorithm, whose expression is shown as:

$$D_{m+1} = \frac{D_m e^{(-\alpha_m y_i h_m(x_i))}}{Z_m} \tag{17}$$

In the above equation, Z_m is the normalization factor and its expression is shown as:

$$Z_m = \sum_{i=1}^N D_m e^{(-\alpha_m y_i h_m(x_i))} \tag{18}$$

Repeating the above steps from Eq.(15) to Eq. (17), we finally obtain a strong classifier whose expression is shown as:

$$G(x) = \text{sign}(\sum_{m=1}^M \alpha_m h_m(x)) \tag{19}$$

The flow of the AdaBoost algorithm is shown in Figure 7. First, the weights of the training samples are initialized and fed into the weak classifier for training. During the training process, the weights of the training samples are continuously updated. The weights of those samples that were misclassified by the previous round of the weak classifier are increased and the weights of those samples that were correctly classified are decreased. In the next training round,

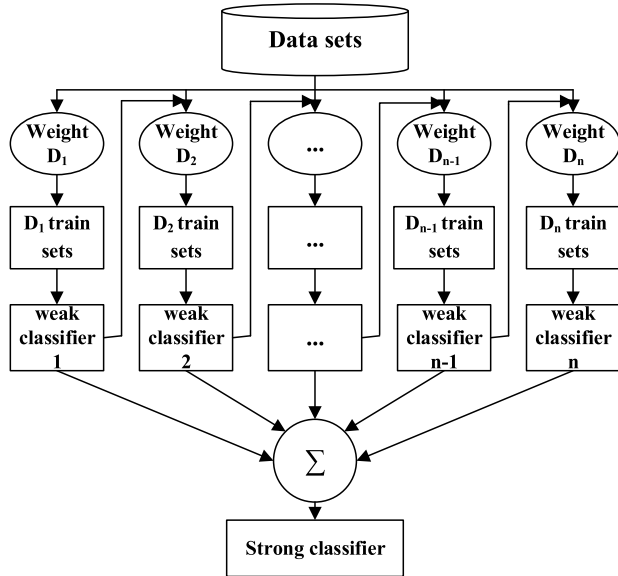


FIGURE 7. Flow of AdaBoost algorithm.

these misclassified samples are given larger training weights, and finally the final output is obtained by weighted voting.

IV. SIMULATION ANALYSIS

In this section, the effectiveness of the above proposed evaluation system will be applied to evaluate and validate the electromagnetic signal recognition effect to illustrate its effectiveness. Specifically, DNN, CNN, AlexNet and ResNet will be used as classifiers to classify 100 classes of signals from WiFi radiation sources, and then the performance of the model will be evaluated comprehensively by the evaluation system. Also, the performance of the AdaBoost algorithm is evaluated.

A. DATASET AND PARAMETER SETTINGS

A WiFi radiation source electromagnetic signal dataset is selected in this paper. The dataset is collected wirelessly in the laboratory and microwave darkroom for LOS, multipath and NLOS multipath channels using FSQ, FSW26 and FSV13 spectrometers to build a rich dataset. The signal acquisition is mainly performed for 100 5GHz WiFi modules for management frames. This contains 100 class I/Q signals. These I/Q signals are stored in the form of a standard binary file.mat file, which is a file format that can be flexibly called by MATLAB, Python and other languages. The training set contains 4000 I/Q signals, each of length 5000. The test set contains 1000 signals, each of length 5000. The data samples in both data sets are labeled correspondingly.

During the training process, the learning rate of each model is set to 0.01.

B. CLASSIFICATION PERFORMANCE EVALUATION OF MODEL ALGORITHMS

In this section, the classification performance of the model and algorithm are tested and evaluated. The model and

algorithm are trained to learn the training set signals and then identify and classify the test set signals. First, the classification performance of four machine learning models is tested by plotting the confusion matrix of different models. Secondly, the classification performance of the AdaBoost algorithm for the test set signals was tested and visualized with the confusion matrix. Finally, the classification performance metrics of the different models and algorithms were calculated and analyzed for comparison.

TABLE 1. Classification performance test results.

Models (Algorithms)	Accuracy	Precision	Recall	F1-score
DNN	77.4%	79.36%	78.00%	77.27%
CNN	84.4%	84.99%	84.91%	84.25%
AlexNet	88.6%	89.82%	89.09%	88.40%
ResNet	94.3%	95.00%	94.76%	94.17%
AdaBoost	85.8%	86.33%	85.95%	85.45%

Table 1 shows the numerical magnitude of the recognition and classification evaluation metrics, including accuracy, precision, recall and F1 score, for the test set by different classification models and AdaBoost algorithms. We can see that for the dataset used in this paper, the ResNet model has the best classification, and the recognition accuracy can reach 94.3%. The DNN model has the lowest recognition accuracy, which can reach more than 77%. The recognition and classification accuracy of CNN can reach more than 84%, and the recognition and classification accuracy of AlexNet can reach more than 88%. For Precision, Recall and F1-score metrics, ResNet has the best recognition and classification results compared to other models and algorithms.

In terms of classification performance, all the above four models have good recognition accuracy. Based on the AdaBoost algorithm for DNN, CNN and AlexNet three base classifiers for comprehensive learning, the signal data set of this experiment is used for training and learning with the base classifiers, and then the final output results are obtained by weighted combination. The experimental results show that its final recognition classification accuracy can also reach 85.8%, but there are also a small number of misclassified samples. Through the analysis of experimental metrics, the AdaBoost algorithm can improve the classification performance of the base classifier to a certain extent, but its classification performance is not as good as its base classifier of AlexNet model recognition. The analysis in this paper suggests that for the AdaBoost algorithm, when its base classifier is a model with strong learning ability, the phenomenon of overfitting may occur during the training process. This phenomenon leads to the insignificant improvement of the classification performance of AdaBoost, and even reduces the recognition accuracy of some base classifiers.

Table 2 shows the results of the classification performance tests for at 0 dB. The test set at 0dB is fed into the model and algorithm for classification and recognition of signals. It can be seen that DNN has the highest recognition accuracy of

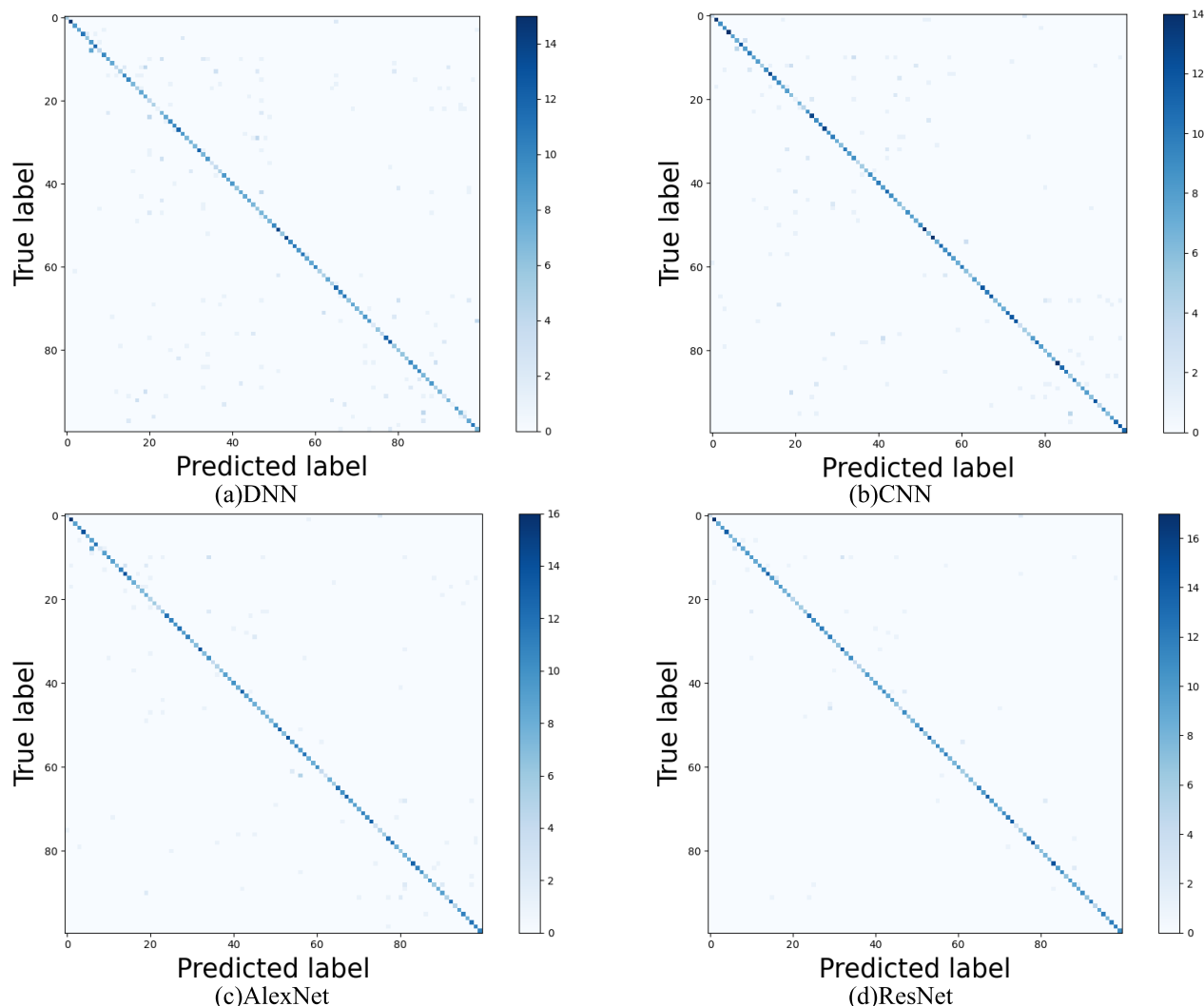


FIGURE 8. Model classification confusion matrix.

TABLE 2. Classification performance test results at 0dB.

Models (Algorithms)	Accuracy	Precision	Recall	F1-score
DNN	76.50%	77.70%	77.65%	76.12%
CNN	31.30%	31.06%	31.07%	26.37%
AlexNet	6.10%	2.53%	6.61%	2.19%
ResNet	11.20%	7.65%	11.71%	6.45%
AdaBoost	44.70%	63.02%	45.09%	44.16%

76.5%, AlexNet has the lowest recognition accuracy of 6.1%. The recognition accuracy of AdaBoost is 44.7%.

Figure 8 shows the confusion matrix of the different classification models for the recognition classification of the test set. For 100 classes of WIFI radiation source signals, all classification models show a straight line on the confusion matrix. DNN can see more misclassified samples on the confusion matrix compared to the other models. CNN and

AlexNet have better classification accuracy than DNN with fewer misclassified samples. ResNet has the best confusion matrix with the least misclassified samples and the clearest straight line presented on the confusion matrix.

Figure 9 shows the confusion matrix of the AdaBoost algorithm in the test set recognition classification. The confusion matrix allows us to visualize the classification performance of the AdaBoost algorithm. In the figure, we can clearly see a clear line of correct classification, with fewer samples being misclassified. Its recognition classification accuracy can reach more than 80%. Compared with the three base classifiers, the classification performance of the algorithm has been improved.

C. COMPLEXITY TEST EVALUATION OF MODELS

In this section, the computational complexity of different models is evaluated and the size of FLOPs is calculated separately for different models and compared analytically.

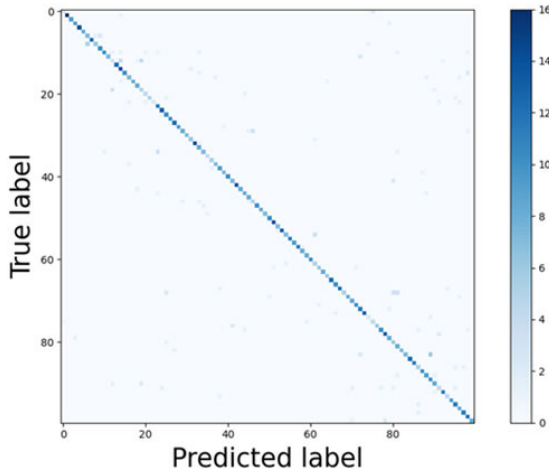


FIGURE 9. AdaBoost algorithm confusion matrix.

TABLE 3. FLOPs test results.

Models	DNN	CNN	AlexNet	ResNet
FLOPs	934283449	80146983	14474273	166623265

Table 3 shows the magnitude of the FLOPs values for different classification models. The ResNet model has large FLOPs and its recognition accuracy is the highest. However, the DNN model has larger FLOPs, but its recognition accuracy is the smallest. The recognition results of the CNN and AlexNet models are similar, but the AlexNet model has smaller FLOPs and better model performance.

The experimental results show that the computational complexity of the model has no direct effect on the model recognition accuracy of this dataset. Among the models with similar recognition classification accuracy, the smaller the complexity, the better the model. Evaluating the models requires considering the issue of model computational complexity based on assessing the classification accuracy of the models.

According to the calculation formula of FLOPs, this index is related to the structure of the model. The index is determined by the number of convolution layers and full connection layers. The FLOPs of the models are not directly related to the robustness of the models as can be seen from the experiments in this paper. The FLOPs of the DNN are the largest and the robustness test shows that the DNN has the best robustness. The ResNet and AlexNet models are less robust, but the FLOPs of the ResNet are larger and the AlexNet is smaller. The FLOPs of AlexNet are smaller.

D. ROBUSTNESS EVALUATION OF MODEL ALGORITHMS

In this section, the robustness of different models and algorithms is evaluated. The recognition classification accuracy of the models and algorithms under different Gaussian

noise disturbances is tested, and a comparative analysis is performed.

Figure 10 shows the recognition accuracy of DNN, CNN, AlexNet, ResNet and AdaBoost algorithms for signals with signal-to-noise ratios of -10 dB to 20 dB. In general, the recognition accuracy of all models or algorithms for this signal dataset gradually improves as SNR increases. The recognition accuracy of ResNet model and AlexNet stabilizes above 16 dB when they tend to. However, when the SNR is below 10 dB, the recognition accuracy of both drops sharply, indicating that the robustness of AlexNet and ResNet models for this electromagnetic signal dataset is poor. DNN is not easily affected by noise, and its recognition accuracy changes less as the SNR decreases. The recognition accuracy of the AdaBoost algorithm decreases flatly when the SNR is below 10 dB.

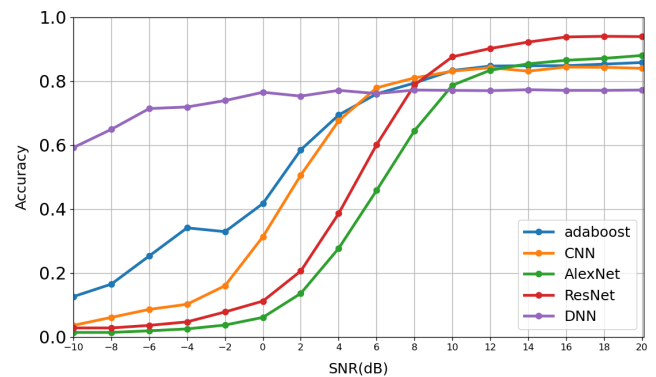


FIGURE 10. Classification performance test under different noise.

Experiments show that noise affects both the machine learning model and the AdaBoost algorithm, with the DNN model having the best robustness and the ResNet model being susceptible to noise. The robustness of the AdaBoost algorithm is better, and the algorithm improves robustness to some extent compared to its weak classifier, but its robustness performance is not higher than the performance of all weak classifiers. Although the DNN model is less effective in recognition and classification, the recognition accuracy of this model is not susceptible to noise. In contrast, the ResNet model has the best recognition performance, but it is most susceptible to noise. This paper argues that this is because the DNN model does not learn the deep features of the samples. For samples with added noise, the model cannot distinguish the effect of noise on such features, so the noise has less impact on the classification performance of the model. On the contrary, the ResNet model learns the deep features of the samples. For the samples with added noise, the model can identify the deep features affected by noise, so the model can identify the wrong samples.

E. EVALUATION TEST SUMMARY

The above experiments evaluate the electromagnetic signal recognition performance of some machine learning models

and algorithms in terms of classification performance, complexity and robustness, respectively. Table 4 shows the relevant experimental results and inferences.

TABLE 4. Summarizing inferences.

Classification Performance	ResNet has the highest recognition accuracy, DNN has the lowest recognition accuracy, and the AdaBoost algorithm does not necessarily improve the recognition accuracy of the base classifier.
Complexity	DNN has the largest FLOPs, which indicates that its computational complexity is the largest. With similar recognition accuracy, the model performance of AlexNet is better because its FLOPs are smaller than that of CNN model. FLOPs are not directly related to the size of classification performance metrics.
Robustness	DNN has the best robustness and ResNet has the worst robustness. The AdaBoost algorithm can improve the robustness of the base classifier to some extent, but its robustness performance is not higher than the performance of all weak classifiers. There is no direct relationship between robustness and classification performance.

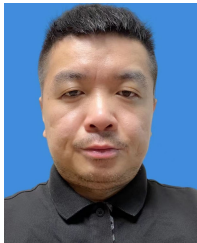
V. CONCLUSION

In this paper, a signal identification and evaluation system based on multidimensional fusion is studied, and the recognition and classification effects of four classification models and AdaBoost algorithm on 100 classes of individual signals from WIFI radiation sources are tested respectively. Through experimental verification, the evaluation system can effectively evaluate the performance of the models and algorithms in three aspects of classification performance, complexity and robustness, and has certain scalability. For this system, more attention can be paid to the evaluation of model security and the evaluation of recognition and classification effects of models and algorithms under certain attacks in the future, so as to form a more comprehensive and integrated evaluation system.

REFERENCES

- [1] S. Zhang, "Electromagnetic signal modulation recognition technology based on lightweight deep neural network," *J. Commun.*, vol. 41, no. 11, pp. 12–21, 2020.
- [2] P. Shao and Z. Chen, "Radio frequency fingerprint feature extraction based on I/Q data distribution features," in *Proc. Photon. Electromagn. Res. Symp. (PIERS)*, Apr. 2022, pp. 161–165.
- [3] H. Wang, J. Li, L. Guo, Z. Dou, Y. Lin, and R. Zhou, "Fractal complexity-based feature extraction algorithm of communication signals," *Fractals*, vol. 25, no. 4, Aug. 2017, Art. no. 1740008.
- [4] Y. Tu, Y. Lin, C. Hou, and S. Mao, "Complex-valued networks for automatic modulation classification," *IEEE Trans. Veh. Technol.*, vol. 69, no. 9, pp. 10085–10089, Sep. 2020, doi: [10.1109/TVT.2020.3005707](https://doi.org/10.1109/TVT.2020.3005707).
- [5] Y. Lin, X. Zhu, Z. Zheng, Z. Dou, and R. Zhou, "The individual identification method of wireless device based on dimensionality reduction and machine learning," *J. Supercomput.*, vol. 75, no. 6, pp. 3010–3027, 2019, doi: [10.1007/s11227-017-2216-2](https://doi.org/10.1007/s11227-017-2216-2).
- [6] T. J. O'Shea, T. Roy, and T. C. Clancy, "Over-the-air deep learning based radio signal classification," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 168–179, Feb. 2018, doi: [10.1109/JSTSP.2018.2797022](https://doi.org/10.1109/JSTSP.2018.2797022).
- [7] Y. Dong, X. Jiang, H. Zhou, Y. Lin, and Q. Shi, "SR2CNN: Zero-shot learning for signal recognition," *IEEE Trans. Signal Process.*, vol. 69, pp. 2316–2329, 2021, doi: [10.1109/TSP.2021.3070186](https://doi.org/10.1109/TSP.2021.3070186).
- [8] C. Hou, G. Liu, Q. Tian, Z. Zhou, L. Hua, and Y. Lin, "Multisignal modulation classification using sliding window detection and complex convolutional network in frequency domain," *IEEE Internet Things J.*, vol. 9, no. 19, pp. 19438–19449, Oct. 2022, doi: [10.1109/IJOT.2022.3167107](https://doi.org/10.1109/IJOT.2022.3167107).
- [9] A. Hirose and S. Yoshida, "Generalization characteristics of complex-valued feedforward neural networks in relation to signal coherence," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 4, pp. 541–551, Apr. 2012, doi: [10.1109/TNNLS.2012.2183613](https://doi.org/10.1109/TNNLS.2012.2183613).
- [10] H. Wang, L. Guo, Z. Dou, and Y. Lin, "A new method of cognitive signal recognition based on hybrid information entropy and D-S evidence theory," *Mobile Netw. Appl.*, vol. 23, no. 4, pp. 677–685, 2018, doi: [10.1007/s11036-018-1000-8](https://doi.org/10.1007/s11036-018-1000-8).
- [11] Y. Lin, Y. Tu, and Z. Dou, "An improved neural network pruning technology for automatic modulation classification in edge devices," *IEEE Trans. Veh. Technol.*, vol. 69, no. 5, pp. 5703–5706, May 2020, doi: [10.1109/TVT.2020.2983143](https://doi.org/10.1109/TVT.2020.2983143).
- [12] Y. Wang, G. Gui, H. Gacanin, T. Ohtsuki, H. Sari, and F. Adachi, "Transfer learning for semi-supervised automatic modulation classification in ZF-MIMO systems," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 10, no. 2, pp. 231–239, Jun. 2020, doi: [10.1109/JETCAS.2020.2992128](https://doi.org/10.1109/JETCAS.2020.2992128).
- [13] Y. Tu, Y. Lin, J. Wang, and J.-U. Kim, "Semi-supervised learning with generative adversarial networks on digital signal modulation classification," *Comput. Mater. Continua*, vol. 55, no. 2, pp. 243–254, 2018, doi: [10.3970/cm.2018.01755](https://doi.org/10.3970/cm.2018.01755).
- [14] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, Aug. 1997.
- [15] J. Guo, P. Ge, W. Jin, and W. Zhang, "Radar signal recognition based on FCBF and AdaBoost algorithm," in *Proc. 37th Chin. Control Conf. (CCC)*, 2018, pp. 4185–4190, doi: [10.23919/ChiCC.2018.8483351](https://doi.org/10.23919/ChiCC.2018.8483351).
- [16] T. Liu, Y. Guan, and Y. Lin, "Research on modulation recognition with ensemble learning," *EURASIP J. Wireless Commun. Netw.*, vol. 2017, no. 1, p. 179, 2017, doi: [10.1186/s13638-017-0949-5](https://doi.org/10.1186/s13638-017-0949-5).
- [17] B. Biggio, Z. Akhtar, G. Fumera, G. L. Marcialis, and F. Roli, "Robustness of multi-modal biometric verification systems under realistic spoofing attacks," in *Proc. Int. Joint Conf. Biometrics (IJCB)*, Oct. 2011, pp. 1–6, doi: [10.1109/IJCB.2011.6117474](https://doi.org/10.1109/IJCB.2011.6117474).
- [18] B. Biggio, G. Fumera, and F. Roli, "Security evaluation of pattern classifiers under attack," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 4, pp. 984–996, Apr. 2014, doi: [10.1109/TKDE.2013.57](https://doi.org/10.1109/TKDE.2013.57).
- [19] Z. Bao, Y. Lin, S. Zhang, Z. Li, and S. Mao, "Threat of adversarial attacks on DL-based IoT device identification," *IEEE Internet Things J.*, vol. 9, no. 11, pp. 9012–9024, Jun. 2022, doi: [10.1109/IJOT.2021.3120197](https://doi.org/10.1109/IJOT.2021.3120197).
- [20] V.-S. Doan, T. Huynh-The, C.-H. Hua, Q.-V. Pham, and D.-S. Kim, "Learning constellation map with deep CNN for accurate modulation recognition," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2020, pp. 1–6, doi: [10.1109/GLOBECOM42002.2020.9348129](https://doi.org/10.1109/GLOBECOM42002.2020.9348129).
- [21] Z. Cai, J. Wang, and M. Ma, "The performance evaluation of big data-driven modulation classification in complex environment," *IEEE Access*, vol. 9, pp. 26313–26322, 2021, doi: [10.1109/ACCESS.2021.3054756](https://doi.org/10.1109/ACCESS.2021.3054756).
- [22] M. Hao and S. Wang, "Evaluation of interference effect of speech communication system," in *Proc. 13th Int. Symp. Antennas, Propag. EM Theory (ISAPE)*, Dec. 2021, pp. 1–3, doi: [10.1109/ISAPE54070.2021.9752950](https://doi.org/10.1109/ISAPE54070.2021.9752950).
- [23] S. Zhang, J. Yu, Z. Bao, S. Mao, and Y. Lin, "Spectrum focused frequency adversarial attacks for automatic modulation classification," 2022, *arXiv:2208.01919*.
- [24] M. Wang, Y. Lin, Q. Tian, and G. Si, "Transfer learning promotes 6G wireless communications: Recent advances and future challenges," *IEEE Trans. Rel.*, vol. 70, no. 2, pp. 790–807, Jun. 2021, doi: [10.1109/TR.2021.3062045](https://doi.org/10.1109/TR.2021.3062045).
- [25] Y. Tu, Y. Lin, H. Zha, J. Zhang, Y. Wang, G. Gui, and S. Mao, "Large-scale real-world radio signal recognition with deep learning," *Chin. J. Aeronaut.*, vol. 35, no. 9, pp. 35–48, Sep. 2022, doi: [10.1016/j.cja.2021.08.016](https://doi.org/10.1016/j.cja.2021.08.016).

- [26] R. Orjesek, R. Jarina, M. Chmulik, and M. Kuba, "DNN based music emotion recognition from raw audio signal," in *Proc. 29th Int. Conf. Radioelektronika (RADIOELEKTRONIKA)*, Apr. 2019, pp. 1–4, doi: 10.1109/RADIOELEK.2019.8733572.
- [27] Y. Wang, M. Liu, J. Yang, and G. Gui, "Data-driven deep learning for automatic modulation recognition in cognitive radios," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 4074–4077, Apr. 2019, doi: 10.1109/TVT.2019.2900460.
- [28] J. Yang and F. Liu, "Modulation recognition using wavelet transform based on AlexNet," in *Proc. IEEE 7th Int. Conf. Comput. Sci. Netw. Technol. (ICCSNT)*, Oct. 2019, pp. 339–342, doi: 10.1109/ICCSNT47585.2019.8962474.
- [29] J. Wang, Y. Zhong, Z. Zheng, A. Ma, and L. Zhang, "RSNet: The search for remote sensing deep neural networks in recognition tasks," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2520–2534, Mar. 2021, doi: 10.1109/TGRS.2020.3001401
- [30] J. Ruan and J. Yin, "Multi-pose face detection using facial features and AdaBoost algorithm," in *Proc. 2nd Int. Workshop Comput. Sci. Eng.*, 2009, pp. 31–34, doi: 10.1109/WCSE.2009.760.



QINGMING LI was born in Changchun, China, in 1978. He received the B.E. and M.E. degrees from the College of Information and Communication Engineering, Harbin Engineering University, Harbin, China. His current research interests include the next generation wireless communication technology, digital signal processing, and high-speed storage technology.



SONGLIN YANG was born in Jilin, China, in 2000. He received the bachelor's degree from the School of Information and Communication Engineering, Harbin Engineering University, China, where he is currently pursuing the master's degree with the School of Information and Communication Engineering. His main research interests include communication, AI security, and model evaluation.



HU CHEN was born in Jingzhou, Hubei, China, in 1991. He received the bachelor's degree from the School of Materials Science and Engineering, Sichuan University, and the doctoral degree from Tsinghua University, China. His current research interests include the machine learning, digital signal processing, and target recognition.

...