

Article

Signal Modulation Recognition Based on DRSLSTM Neural Network

Ping Tan ^{1,2} , Dongxu Chen ¹, Kaijun Zhou ^{1,2,*} , Yi Shen ^{1,2,*}  and Shen Zhao ¹

¹ School of Intelligent Engineering and Intelligent Manufacturing, Hunan University of Technology and Business, Changsha 410205, China; tp2008@hutb.edu.cn (P.T.); cdx909136606@163.com (D.C.); zhaoshen@hutb.edu.cn (S.Z.)

² Xiangjiang Laboratory, Changsha 410205, China

* Correspondence: zkj@hutb.edu.cn (K.Z.); shenyi@hutb.edu.cn (Y.S.); Tel.: +86-139-7317-2416 (K.Z.); +86-151-5069-8138 (Y.S.)

Abstract

To overcome the challenge of degraded classification accuracy in automatic modulation recognition under low signal-to-noise ratio (SNR) conditions, this paper introduces an end-to-end framework utilizing a Deep Residual Shrinkage Long Short-Term Memory (DRSLSTM) network. The proposed DRSLSTM architecture synergistically integrates two dedicated components: a deep residual shrinkage module, specifically designed for I/Q signals to perform simultaneous denoising and spatial feature extraction, and a Long Short-Term Memory (LSTM) network that captures long-range temporal dependencies from the refined feature sequences. Extensive simulations on a public dataset show that the DRSLSTM model achieves a recognition accuracy of 51.19% at -8 dB SNR, an improvement of 3.36 percentage points over the CLDNN baseline, and consistently surpasses six benchmark models at SNR levels above 0 dB. Moreover, it exhibits higher average recognition accuracy across a wide range of SNR conditions. The experimental results validate the overall superiority of the proposed approach.

Keywords: modulation recognition; low SNR; DRSLSTM neural network



Academic Editor: Alberto Fernandez Hilario

Received: 29 September 2025

Revised: 2 November 2025

Accepted: 11 November 2025

Published: 13 November 2025

Citation: Tan, P.; Chen, D.; Zhou, K.; Shen, Y.; Zhao, S. Signal Modulation Recognition Based on DRSLSTM Neural Network. *Electronics* **2025**, *14*, 4424. <https://doi.org/10.3390/electronics14224424>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the widespread deployment of modern wireless communication technologies, communication devices have become pervasive in everyday life, resulting in increasingly complex electromagnetic environments and congested radio spectrum resources [1–3]. As a fundamental capability for radio spectrum sensing, automatic modulation recognition (AMR) plays a critical role in enabling efficient spectrum management [4].

Automatic modulation recognition methods are conventionally categorized into three principal paradigms: likelihood ratio-based approaches, feature extraction-based techniques, and deep learning-based methods [5]. Likelihood ratio-based methods perform classification by applying the maximum likelihood criterion to the received signals, with their theoretical performance bound contingent upon the availability of an accurate channel model. However, in non-cooperative communication scenarios, precise channel models are frequently unavailable. This model mismatch often leads to significant performance degradation under complex or unpredictable channel conditions, thereby limiting the practical utility of these methods in such environments.

Feature extraction-based methods rely on the time-domain [6], frequency-domain [7], or statistical characteristics of signals. Yet, under low SNR conditions, these features fail to adequately represent signal properties, resulting in diminished recognition accuracy [8–10].

In recent years, deep learning has achieved remarkable success in domains such as speech and image recognition [11,12]. Its application has extended to signal recognition, offering a new paradigm for AMR [13,14]. Deep learning enables end-to-end learning, allowing features to be extracted directly from raw signals, without the need for manual feature engineering or preprocessing steps. This not only simplifies the recognition pipeline but also enhances the stability and reliability of performance. In scenarios involving numerous modulation types and low SNR, deep learning models can capture subtle signal characteristics through large-scale data training, thereby improving both accuracy and robustness.

Subsequent research has progressively advanced the application of deep learning [15] in this domain. Pioneering work by O'Shea et al. established early benchmarks using GNU Radio-generated datasets, demonstrating the feasibility of end-to-end modulation recognition with deep neural networks [16,17]. To enhance model generalizability, further work expanded the dataset to incorporate a wider range of signal types and realistic channel impairments. Results, however, highlighted persistent challenges under low SNR, as recognition rates remained below 60% at 0 dB SNR [18,19].

More recent efforts have explored diverse neural architectures to improve robustness [20]. Enhanced digital modulation recognition was demonstrated in [21,22] through the integration of attention mechanisms with a ResNet50 backbone. And LSTM neural network, renowned for their ability to model long-range temporal dependencies in sequential data, have also been applied to modulation recognition [23]. However, their effectiveness on MPSK signals was limited by insufficient spatial feature extraction. More recently, promising gains in accuracy have been achieved through the adoption of Transformer-based models [24].

Despite considerable progress in the field, achieving accurate recognition under low SNR conditions remains a critical challenge, primarily due to limited noise robustness and the disjointed learning of spatiotemporal features [25,26]. This paper addresses this gap by proposing a novel Deep Residual Shrinkage Long Short-Term Memory (DRSLSTM) network model. The model's principal innovation is the synergistic combination of a deep residual shrinkage module for concurrent noise reduction and spatial feature extraction, and an LSTM network for temporal modeling. This integrated end-to-end architecture is specifically designed to enhance discrimination capability in low SNR environments, demonstrating superior performance compared to existing methods.

2. Related Works

2.1. Modulation Technologies

Modulation technology serves as a fundamental component of wireless communication systems, enabling the mapping of baseband information signals to carrier waveforms suitable for transmission over wireless channels. Based on the variation of carrier parameters, such as amplitude, phase, and frequency, common modulation schemes can be classified into several types, including amplitude-shift keying (ASK), phase-shift keying (PSK), frequency-shift keying (FSK), amplitude modulation (AM), and frequency modulation (FM). The principles, key characteristics, and typical application scenarios of these schemes are elaborated in the following:

2.1.1. ASK Technologies

ASK encodes information by varying the amplitude of the carrier signal, while maintaining constant frequency and phase. It is one of the earliest and simplest digital modulation schemes, and its typical variants include On–Off Keying (OOK), 4ASK, and 8ASK.

OOK is the most basic form of ASK, where the carrier is “on” (transmitted) to represent the digital symbol “1” and “off” (not transmitted) to represent “0”. 4ASK (4-level ASK) and 8ASK (8-level ASK) are multi-level extensions of OOK. In 4ASK, the carrier amplitude is divided into 4 discrete levels, each representing 2 bits of digital information (e.g., “00”, “01”, “10”, “11”). 8ASK further increases the amplitude levels to 8, enabling each level to carry 3 bits of information.

The advantage of multi-level ASK (4ASK, 8ASK) lies in its higher spectral efficiency compared to OOK, where under the same bandwidth, it can transmit more data. However, the increase in amplitude levels also reduces the distance between adjacent signal points, making 4ASK and 8ASK more vulnerable to noise interference, thus requiring higher SNR conditions for reliable communication. They are commonly used in high-speed data transmission systems such as digital microwave communication.

2.1.2. PSK Technologies

PSK encodes digital data by varying the phase of the carrier wave while maintaining a constant amplitude. This method exhibits strong noise immunity due to its constant-envelope characteristic, making it particularly suitable for high-reliability wireless communication systems. Commonly employed variants include Binary Phase-Shift Keying (BPSK), Quadrature Phase-Shift Keying (QPSK), and 8PSK.

BPSK is the fundamental PSK scheme, which uses two opposite carrier phases (e.g., 0° and 180°) to represent the digital symbols “0” and “1”. QPSK uses four discrete carrier phases (e.g., 0° , 90° , 180° , 270°) to represent digital information, with each phase corresponding to 2 bits of data [27,28]. 8PSK extends the phase levels of QPSK to 8, with each phase representing 3 bits of digital information. While this higher-order configuration achieves a spectral efficiency of up to 3 bits/s/Hz, surpassing both BPSK and QPSK, the reduced Euclidean distance between adjacent constellation points renders 8PSK more vulnerable to phase noise and channel fading impairments.

2.1.3. Amplitude Modulation (AM) and Frequency Modulation (FM)

AM and FM are traditional analog modulation technologies, but they still play important roles in specific communication scenarios due to their mature implementation and compatibility.

AM-SSB-WC (Amplitude Modulation-Single Sideband-Wideband Carrier) is a variant of amplitude modulation. In conventional double-sideband AM (DSB-AM), the modulated signal contains the carrier and two sidebands (upper and lower), which waste bandwidth. Single-sideband AM (SSB-AM) suppresses one sideband and retains the other, significantly reducing the bandwidth occupation (by half compared to DSB-AM).

FM modulates the frequency of the carrier signal according to the variation of the baseband information signal, while the carrier amplitude remains constant [29]. The key advantage of FM is its strong anti-noise performance—since noise primarily affects signal amplitude (which can be suppressed by amplitude limiters in the receiver), FM is less affected by noise compared to AM.

However, FM generally exhibits lower spectral efficiency than modern digital modulation schemes such as QPSK and is more susceptible to performance degradation under Doppler frequency shifts, which restricts its use in high-speed mobile communication environments.

2.2. Long Short-Term Memory (LSTM) Neural Networks

The Long Short-Term Memory (LSTM) network is a pivotal variant of recurrent neural networks (RNNs) [30], specifically designed to overcome the vanishing and exploding gradient problems inherent in conventional RNN architectures [31]. By incorporating gating mechanisms, LSTM models are capable of effectively capturing long-range temporal dependencies in sequential data. This ability has established LSTM as a fundamental building block in a variety of time-series-processing applications, including natural language processing (NLP), speech recognition, and communication signal modulation recognition.

The key innovation of LSTM lies in its specialized memory cell structure, which enables the network to dynamically “remember” or “forget” information over long sequences through three gating mechanisms: the forget gate, input gate, and output gate. Each gate is composed of a sigmoid activation function that controls the flow of information, while a tanh activation function regulates the magnitude of the information stored in the memory cell.

2.2.1. Memory Cell

The memory cell (C_t) functions as the central component of the LSTM unit, tasked with maintaining long-term temporal dependencies. Its state is iteratively updated using the current input (x_t) and the previous hidden state (h_{t-1}), regulated by the coordinated operation of the three gating mechanisms. This gated architecture enables the LSTM to retain essential historical information, such as long-term trends in signal amplitude or phase, while filtering out irrelevant noise. As a result, the model is particularly suitable for analyzing non-stationary time-series data, including modulated communication signals.

2.2.2. Gating Mechanisms

Forget Gate (f_t): This gate controls the extent to which information from the previous memory cell state (C_{t-1}) is preserved. It is computed as Equation (1):

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

where σ denotes the sigmoid function, W_f and b_f are the weight matrix and bias vector of the forget gate, respectively, and $[h_{t-1}, x_t]$ represents the concatenation of the previous hidden state and current input. A gate output near 1 implies that the corresponding information is largely retained, whereas a value near 0 indicates that it is suppressed.

Input Gate (i_t): The input gate regulates the incorporation of new information into the memory cell by evaluating the current input x_t and the previous hidden state h_{t-1} . It operates through two synergistic components: a sigmoid-activated layer that determines the extent of state update, and a tanh-activated layer that generates a candidate cell state ($\tilde{C}_t$). The formulas are Equations (2) and (3):

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (3)$$

where W_i , W_C , b_i , and b_C are the weights and biases of the input gate and candidate layer. The updated memory cell state is then calculated as Equation (4):

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (4)$$

where $\odot$ denotes the element-wise multiplication operation.

Output Gate (o_t): This gate determines the hidden state (h_t) output by the current LSTM unit, which is used as the input for the next time step and the final output of the

network. It first uses a sigmoid layer to select the information from the memory cell, then passes the memory cell state through a tanh layer (to scale values to $[-1, 1]$) and multiplies it by the sigmoid output to generate h_t . The formulas are Equations (5) and (6):

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t \odot \tanh(C_t) \quad (6)$$

where W_o and b_o are the weight matrix and bias vector of the output gate.

Compared to traditional RNNs and other time-series models, LSTM exhibits two key advantages for processing sequential data such as communication signals:

- (a) **Long-Term Dependency Capture:** By regulating the flow of information through gating mechanisms, LSTM effectively mitigates the vanishing gradient problem. For example, in signal modulation recognition, where the modulation mode of a signal is determined by the temporal correlation of its I/Q components over multiple time steps, LSTM can retain the phase and amplitude trends of the signal over hundreds of time steps, whereas traditional RNNs would lose this information due to gradient decay.
- (b) **Robustness to Noise:** The forget gate of LSTM can adaptively discard noise components in the sequence by assigning low weights to irrelevant fluctuations. This inherent noise suppression capability makes LSTM more suitable for low SNR scenarios compared to shallow feature extraction methods.

3. DRSLSTM Neural Network

3.1. DRSLSTM Model

The Deep Residual Shrinkage Long Short-Term Memory (DRSLSTM) neural network architecture, which integrates residual shrinkage blocks with long short-term memory networks, is designed to perform robust automatic modulation recognition from raw I/Q signals. The overall network structure is illustrated in Figure 1. After dimension expansion, the I/Q data undergoes a Conv2D layer, a Batch Normalization (BN) layer, a Rectified Linear Unit (ReLU) activation function, and a final MaxPooling2D layer to extract spatial features. These extracted features are fed into the parallel branches.

In one branch, the features are input to a Residual Shrinkage Unit (RSU). First, two convolutional layers are used for deeper feature extraction. The absolute values of the extracted features are computed and then processed by Global Average Pooling (GAP).

In the other branch of the RSU, the original features are passed through a Dense layer and a BN layer, followed by scaling via a Sigmoid function to generate thresholds corresponding to the features. These thresholds are multiplied with the GAP-processed features, and the product is subjected to soft thresholding to output denoised spatial features of the signal.

Subsequently, the denoised spatial features are fed into six consecutive RSU modules for further feature refinement, and then input to an LSTM layer to extract temporal features. This step is designed to capture long-term dependencies between signal features, thereby obtaining temporal characteristics of the signal. The extracted temporal features are then passed to a Dense layer (Dense1) with 128 neurons. To prevent overfitting and ensure the model's generalization ability, a dropout layer is incorporated, with a random dropout rate set to 0.5. Finally, the classification results are output through a Softmax activation function.

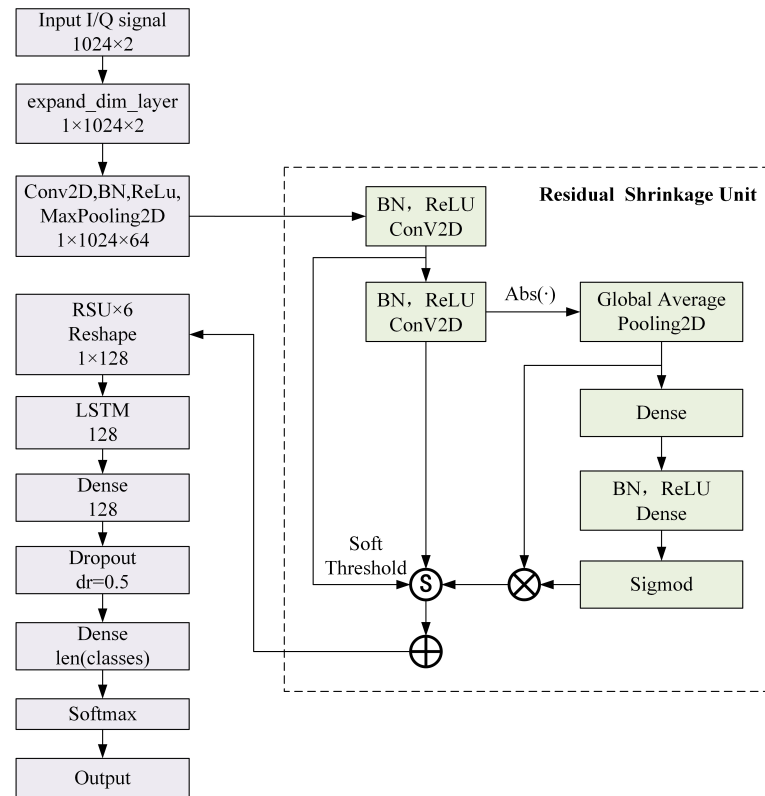


Figure 1. Structure Diagram of DRSLSTM Neural Network. Note: Abs(·) is the absolute value of input.

3.2. Residual Shrinkage Unit

During signal transmission, the integrity of waveforms is often compromised by channel impairments such as additive noise and multipath propagation, resulting in discrepancies between the transmitted and received signals. These distortions hinder the network's ability to accurately characterize underlying modulation patterns. The proposed DRSLSTM architecture addresses this issue by employing convolutional layers to extract discriminative spatial features and LSTM layers to model temporal dependencies within the sequence. Furthermore, the RSUs are incorporated to automatically focus on important features and suppress noise during spatial feature extraction.

The RSU integrates a soft thresholding function, an attention mechanism, and a deep residual network (ResNet). A key advantage of this design is its use of the attention mechanism to adaptively set thresholds, enabling the soft thresholding function to suppress noise-related components while preserving critical features. As a result, the RSU enhances the capability of deep neural networks to learn discriminative representations from noisy signal data.

The soft-thresholding technique serves as a well-established approach for enhancing signal discriminability by suppressing noise-like components whose absolute values fall below a predefined threshold. The soft-thresholding formula is shown in Equation (7):

$$y = \begin{cases} x - \tau & x > \tau \\ 0 & -\tau \leq x \leq \tau \\ x + \tau & x < -\tau \end{cases} \quad (7)$$

The soft-thresholding function determines whether to retain the input value x based on the set threshold τ . Only when the absolute value of the input exceeds the threshold will the function retain the corresponding value; otherwise, it will be set to zero within the threshold range. For each sample, the noise content varies. Setting different thresholds for

denoising each sample can achieve a more effective denoising effect, ensuring the flexibility and accuracy of processing among different samples.

In the RSU, the specific realization of this process is that signal features pass through a small attention sub-network, which can learn a set of independent thresholds, and then perform soft-thresholding on each feature channel. As shown in Equation (8), in this network structure, the absolute value of the input feature x is first taken, and then global average pooling (GAP) is performed to obtain the feature W .

$$W = \text{average}|x_{i,j,c}| \quad (8)$$

The feature vector obtained from the GAP operation is fed into an attention sub-network. This sub-network processes the input through two fully connected layers followed by a Sigmoid activation function, generating a set of channel-wise scaling coefficients α_c between 0 and 1. The computation of α_c is defined as follows:

$$\alpha_c = \frac{1}{1 + \exp(-z_c)} \quad (9)$$

where z_c is the feature of the c -th neuron, and α_c is the c -th scaling parameter processed by the Sigmoid function. As shown in formula Equation (10), the threshold τ_c can be expressed as the product of α_c and the feature W .

$$\tau_c = \alpha_c \times W \quad (10)$$

In the DRSLSTM model, the soft-thresholding function is employed to selectively process signal features: components with magnitudes exceeding a predefined threshold are preserved, while those below the threshold are attenuated toward zero. This mechanism effectively suppresses noise-dominated elements and enhances the discriminative features within the signal. As a result, the model achieves improved automatic modulation recognition accuracy under low SNR conditions by refining the feature representation and increasing robustness to noise interference.

4. Experiments and Results

4.1. Dataset and Experimental Environment

The public dataset RML2018.01a is employed for training and evaluating the proposed DRSLSTM model. This dataset contains radio signals represented as I/Q samples, structured in a 1024×2 matrix format to capture the temporal evolution of the signal's complex envelope. In this experiments, In this experiment, eight modulation schemes were selected, with the SNR configured to range from -8 dB to 10 dB in increments of 2 dB, yielding a total of $327,680$ data samples. The detailed parameters of the dataset are presented in Table 1.

The experimental setup for training and evaluating the DRSLSTM model was implemented on the Google Colaboratory (Colab) cloud platform. The hardware acceleration was provided by an NVIDIA L4 GPU, leveraging CUDA architecture to expedite computational operations during both training and inference phases.

The initial values of the key hyperparameters for our model, including the learning rate, batch size, and number of training epochs, were primarily chosen based on established practices in related research concerning model architectures. Subsequently, a primary screening was conducted, focusing on the learning rate due to its well-documented critical impact on model convergence and final performance. We performed a grid search over a predefined range of values which are 0.1 , 0.01 , 0.001 , and 0.0001 . The learning rate of 0.001 was ultimately selected as it yielded the most stable convergence and the highest validation

accuracy during our experiments, striking an optimal balance between training speed and final performance. The principal hyperparameters configured for the DRSLSTM model are summarized in Table 2.

Table 1. Detailed Parameters of Dataset.

Dataset Content	Parameter
Modulation Types	'OOK', '4ASK', '8ASK', 'BPSK', 'QPSK', '8PSK', 'AM-SSB-WC', 'FM'
SNR Range	−8 dB:2 dB:10 dB
Number of Samples	327,680
Sampling Frequency	1 MHz
Sample Format	1024 × 2 I/Q data
Roll-off Factor	0.35
Number of Samples per Symbol	8
Maximum Carrier Offset and Its Standard Deviation	500 Hz, 0.01 Hz
Channel Fading Model	Rayleigh Fading

Table 2. Table of Network Training Parameters.

Network Training Hyperparameters	Value
Maximum Training Iterations	200
Batch Size	512
Initial Learning Rate	0.001
Learning Rate Decay Factor	0.5
Learning Rate Decay Patience Epochs	4
Optimizer	Adam
Loss Function	CCE Loss

4.2. Results of DRSLSTM Model

A confusion matrix serves as a fundamental performance evaluation tool for classification models, providing a systematic breakdown of prediction outcomes against actual class labels. Each cell in this matrix contains the count or proportion of samples corresponding to a specific combination of the true class and the predicted class, thereby offering a detailed view of the model's discriminative capability.

Figure 2 is the average confusion matrix under SNRs of −8~10 dB, which shows the comparison between the predicted values and true values of the DRSLSTM model for 8 modulation schemes. As presented in the figure, the recognition rates of BPSK, AM-SSB-WC, and FM all reach 95% or higher, with a very low probability of misclassification into other modulation schemes. The recognition rate of 4ASK reaches 75%, that of 8ASK reaches 74%, and that of QPSK reaches 76%. For the remaining modulation schemes, the recognition rate of OOK reaches 89%, and that of 8PSK reaches 83%.

Regarding modulation schemes within the same category, the recognition rate of OOK among MASK-type modulations reaches 89%, while BPSK achieves the highest accuracy of 97% within the MPSK family. This pattern indicates that modulation schemes with lower orders generally exhibit higher intra-category discriminability, as simpler constellations possess more distinct inter-symbol distances, reducing the likelihood of misclassification under noisy conditions.

Furthermore, the confusion matrices reveal that most misclassifications for MASK and MPSK categories occur within their respective modulation families, with only a minor proportion of samples being incorrectly assigned to other categories. For instance, 15% of 4ASK signals are misclassified as 8ASK, whereas merely 5% are confused with AM-SSB-WC, underscoring the structural similarity among higher-order variants within the same modulation class.

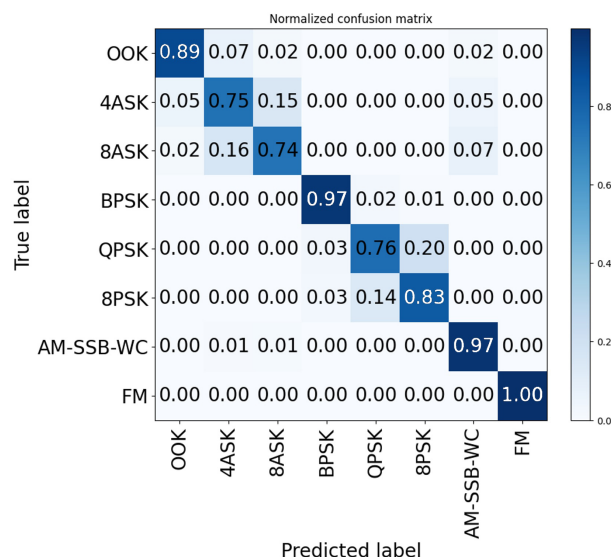


Figure 2. Average Confusion Matrix of DRSLSTM under SNRs of $-8\sim 10$ dB.

To comprehensively assess the model's stability and mitigate random error risks, we conducted 5-fold cross-validation repeated 5 times. This approach involves partitioning the dataset into 5 folds, performing the training and validation process 5 separate times with different random splits, and ultimately calculating the average performance metrics along with their standard deviations across all runs. The results, detailed in the updated manuscript, demonstrate that our model achieves an average accuracy of $86.30 \pm 0.48\%$ (mean $\pm$ standard deviation) across all cross-validation folds.

Overall, the DRSLSTM model achieves good recognition performance for these 8 modulation schemes. This indicates that when processing complex communication signals, the model can not only effectively address the challenges under low SNR conditions but also achieve effective distinction among multiple modulation schemes. In addition, the recognition rates of 4ASK, 8ASK, and QPSK are relatively lower than those of other modulation schemes, which also reflects that the model has a certain degree of error when processing similar modulation schemes.

In Figure 3, the confusion matrices of the DRSLSTM network are listed under SNRs of -8 dB, -6 dB, -4 dB, -2 dB, 0 , 2 dB, 4 dB, 6 dB and 8 dB. It can be observed that as the SNR increases, the color on the diagonal of the confusion matrices becomes darker, which indicates a higher recognition accuracy and better performance in automatic modulation recognition.

When the SNR is -4 dB, the recognition rates of the two analog modulation signals (AM-SSB-WC and FM) have basically reached 100%, which are hardly affected by noise; at this SNR, the recognition rate of BPSK has also reached 100%. When the SNR is 0 dB, the recognition rate of OOK has also achieved 100%. At an SNR of 2 dB, the recognition rates of QPSK and 8PSK are close to 100%. When the SNR is 8 dB, the recognition rate of 4ASK has basically reached 100%.

As the SNR varies from -6 dB to -2 dB, the recognition accuracy for QPSK modulation exhibits a non-monotonic trend, initially decreasing from 0.42 to 0.20 and then returning to 0.42 . Analysis of the confusion matrix reveals that the misclassification primarily occurs between QPSK and 8PSK. Specifically, at the lowest accuracy point (SNR = -4 dB), approximately 20% of the QPSK signals are correctly identified, while the remaining 80% are misclassified as 8PSK. It is hypothesized that the dominant cause of this phenomenon is the significant phase deviation induced by intense noise at low SNR levels. At around -4 dB, the noise intensity causes the signal characteristics to shift closer to the decision boundary region of 8PSK modulation, leading to frequent misclassification. This interpretation is

further supported by the observed reciprocal misclassification at -8 dB, where 8PSK signals are misidentified as QPSK with a probability of 0.65. These mutual misclassifications indicate that modulation schemes within the same family (such as PSK-type modulations) are prone to confusion under low-SNR conditions due to their susceptibility to similar noise-induced distortions, particularly phase ambiguity.

As the SNR increases, the modulation recognition performance of the DRSLSTM network improves markedly, a trend clearly reflected in the confusion matrices. At SNR levels of 6 dB and 8 dB, the model achieves high classification accuracy across most modulation types, demonstrating its effectiveness in discriminating between diverse signal formats while maintaining robust performance under moderate noise conditions.

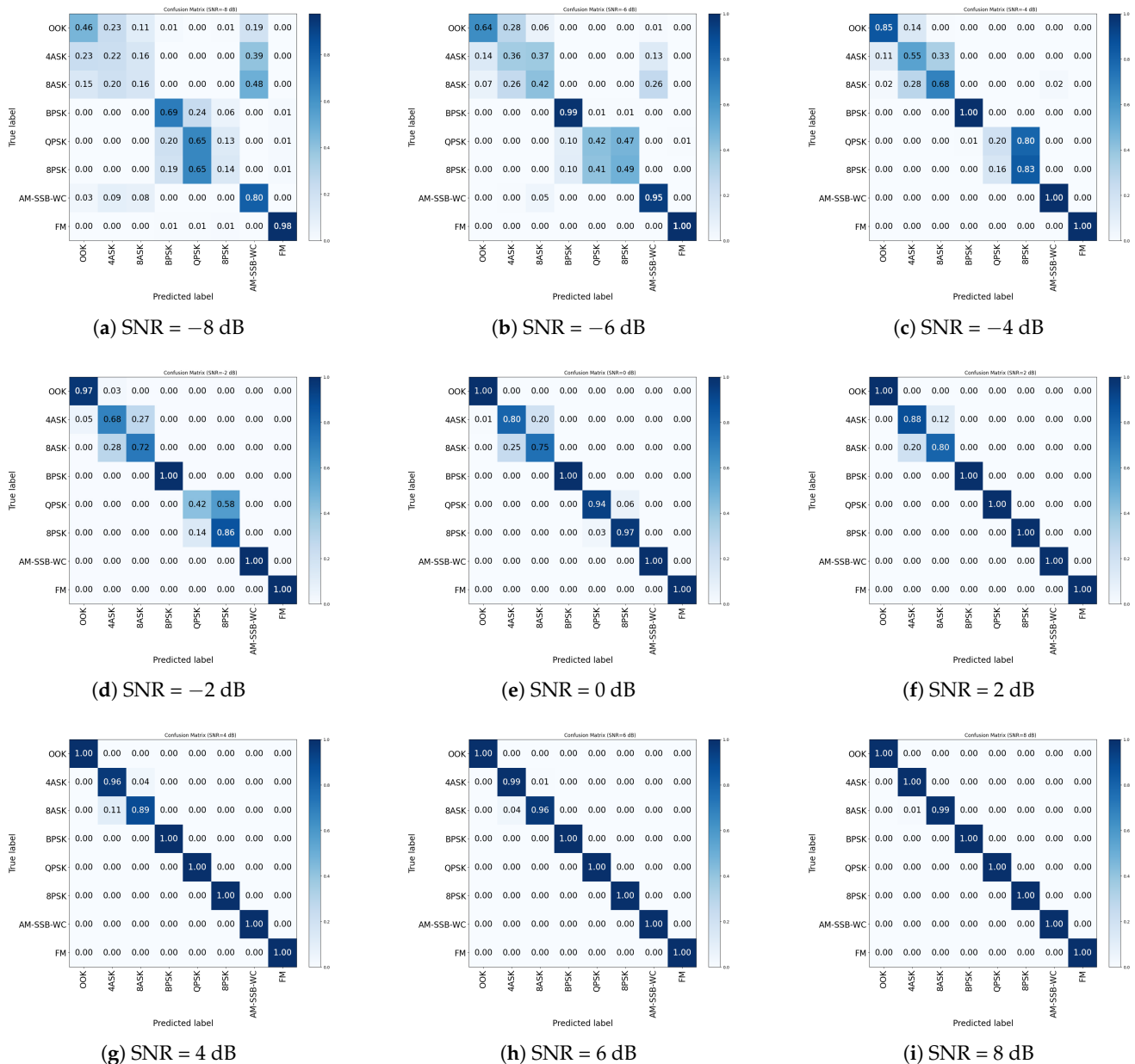


Figure 3. Confusion matrices for the proposed DRSLSTM on the RML2018.01a dataset at -8 dB, -6 dB, -4 dB, -2 dB, 0, 2 dB, 4 dB, 6 dB, 8 dB SNR.

4.3. Comparison with Other Models

To comprehensively assess the recognition performance of the DRSLSTM model under varying SNR conditions, this section presents a comparative analysis between the

proposed DRSLSTM and six benchmark models: ResNet [21], CNN [26], CLDNN (Convolutional Long Short-Term Deep Neural Network) [32], LSTM [23], Transformer [24], and TLDNN [33]. These six models cover representative architectures across distinct technical paradigms in deep learning, ensuring the comparative analysis is both comprehensive and representative. Specifically, CNN and ResNet excel in capturing spatial hierarchies and fine-grained local patterns; LSTM is specialized in modeling temporal dependencies and long-range contextual information within sequential data; CLDNN integrates the advantages of convolutional feature extraction and recurrent sequence modeling, enabling it to handle complex spatiotemporal data effectively; Transformer is proficient in global dependency capture, with its self-attention mechanism facilitating efficient modeling of long-range interdependencies across entire data sequences; and TLDNN, as a recently proposed novel method, is included to incorporate the latest state-of-the-art advancements in the field.

The evaluation employs the same dataset described in Section 4.1, with all experiments conducted under identical hyperparameter settings to ensure a fair comparison. Table 3 summarizes the comparative performance of the six benchmark models, reporting their average recognition accuracy across SNR conditions ranging from -8 dB to 10 dB. These results enable a systematic examination of each model's robustness and discriminative capability under varying noise levels.

Table 3. Modulation Recognition Rates of 7 Methods.

SNR (dB)	CLDNN	Resnet	LSTM	CNN	Transformer	TLDNN	DRSLSTM
-8	47.83%	44.23%	37.75%	48.53%	38.91%	47.37%	51.19%
-6	61.93%	58.83%	50.65%	63.15%	56.87%	64.54%	65.76%
-4	75.05%	74.51%	65.66%	74.65%	73.98%	75.51%	76.44%
-2	81.83%	82.46%	74.81%	79.70%	85.58%	82.58%	83.16%
0	90.66%	92.48%	80.94%	88.89%	93.35%	92.65%	93.10%
2	95.05%	95.72%	89.63%	95.45%	95.57%	94.72%	95.99%
4	97.31%	97.45%	93.51%	97.11%	96.68%	97.19%	98.18%
6	98.36%	98.48%	95.06%	97.78%	97.12%	98.64%	99.45%
8	99.10%	98.92%	95.46%	98.04%	97.49%	99.36%	99.86%
10	99.22%	99.10%	95.27%	98.47%	97.64%	99.75%	99.86%
Average	84.63%	84.20%	77.87%	84.18%	83.31%	85.23%	86.30%

Note: Bold formatting indicates the best results.

It can be seen that as the SNR increases, the performance of the six models improves significantly. Especially when the SNR is 0 dB or above, the accuracy generally exceeds 90% . When the SNR ranges from -8 dB to 10 dB, the DRSLSTM generally achieves the best comprehensive performance, except that it is slightly surpassed by Transformer at -2 dB and 0 dB. At 10 dB, all models achieve very high accuracy. The accuracy of DRSLSTM, CLDNN, and Resnet is close to 100% , while that of LSTM, CNN, and Transformer is also above 95% .

Compared with the LSTM model, the DRSLSTM model achieves an 8.43% improvement in average recognition rate and maintains a higher recognition rate under all SNR conditions. Particularly under low SNRs (below 0 dB), the average recognition rate of the DRSLSTM model is more than 10% higher than that of the LSTM. Additionally, among the 6 network models, the LSTM exhibits the lowest recognition rate under each SNR. This result indicates that relying solely on extracting temporal features of modulated signals has limited effectiveness in enhancing recognition accuracy.

Comparison with the CLDNN model, the DRSLSTM model shows a 1.67% increase in average recognition rate. Under low SNRs ranging from -8 dB to 0 dB, its recognition rate is approximately 3% higher than that of the CLDNN. When the SNR is between 2 dB and

10 dB, although the DRSLSTM still has a higher recognition rate than the CLDNN, the gap is minimal, and their recognition rates are relatively close.

Similarly, under low SNRs (−8 dB to 0 dB), the DRSLSTM model outperforms the CNN, ResNet and TLDNN models in recognition rate. Specifically, at −8 dB, the recognition rate of the DRSLSTM network is 6.96% higher than that of Resnet; at 0 dB, it is 4.21% higher than that of CNN. When the SNR exceeds 2 dB, the DRSLSTM network still maintains a higher signal recognition rate than CNN and Resnet.

Under low SNRs, the DRSLSTM model also outperforms the Transformer. For example, at −8 dB, the recognition rate of the Transformer is only 38.91%, while that of the DRSLSTM network reaches 51.19%. However, the Transformer surpasses the DRSLSTM network at 0 dB and its recognition rate plateaus at around 97% when the SNR exceeds 6 dB. This leads to an approximately 3% lower average recognition rate of the Transformer compared with the DRSLSTM.

The above comparative analysis can be shown in the Figure 4, which indicates that that under high SNR conditions, where the signal is relatively clean and noise interference is minimal, the noise suppression capability of the network is less utilized, resulting in performance levels comparable to those of other benchmark models. In contrast, under low-SNR scenarios, the DRSLSTM effectively leverages its residual shrinkage mechanism to suppress noise interference and extract discriminative features directly from the raw data, thereby achieving superior recognition accuracy compared to other models.

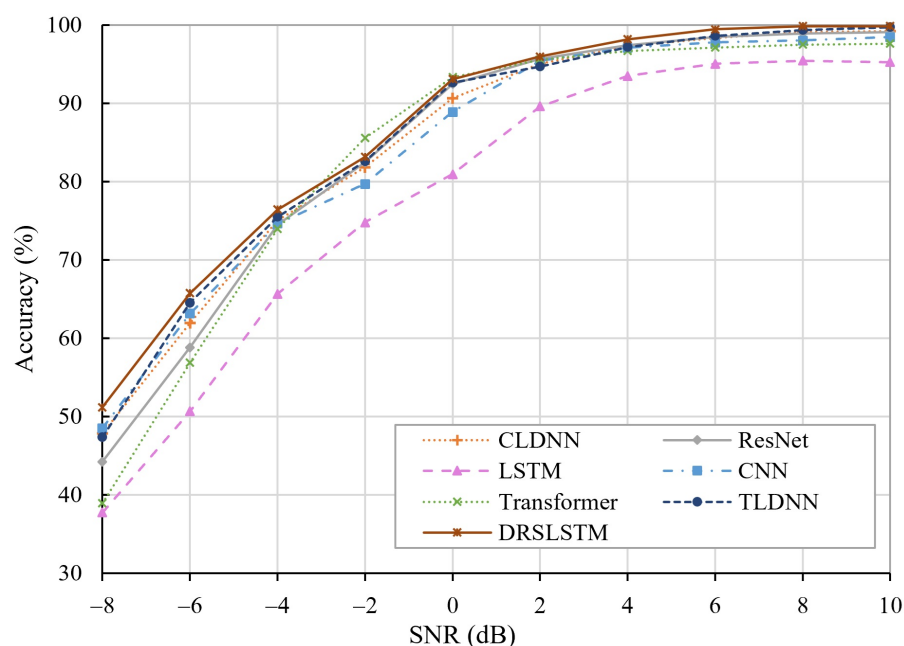


Figure 4. Recognition accuracy comparison of different methods across SNR ranges on the RML2018.01a dataset.

The experimental results establish that Automatic Modulation Recognition, the classification of modulation schemes from raw I/Q sequences, requires the joint capture of local spectral features and long-term temporal dependencies. In these benchmark methods, CNN excels at extracting local, translation-invariant patterns (e.g., pulse shapes, sudden phase shifts) but struggle with long-range temporal relationships crucial for identifying modulation. LSTM is adept at modeling the long-term temporal dynamics (e.g., phase trajectories over time). However, its sequential processing is slow and it may underutilize local hierarchical features. ResNet’s deep residual learning enables stable training of very deep networks, effectively capturing complex, hierarchical feature representations from

I/Q data, though it lacks an explicit, native mechanism for temporal modeling. Transformers theoretically excel at capturing global dependencies across the entire signal sequence via self-attention. However, their quadratic complexity is a significant limitation for long radio signal captures, and they often require substantial data to generalize effectively in this domain. Hybrid Models (CLDNN/TLDNN) synergize these strengths: CNNs extract discriminative local features, which LSTMs then process temporally. This makes them powerful baselines for AMR, as they address both key requirements. In a similar vein, DRSLSTM integrates deep residual shrinkage module connections with LSTM, a hybrid design that yields excellent performance in low-noise environments.

To further validate the generalizability of the proposed algorithm, DRSLSTM was subjected to an additional systematic evaluation against 6 methods utilizing the RML2016.10a dataset. Table 4 presents the aggregated average performance metrics for all algorithms, consolidating the results obtained from both the RML2016.10a and RML2018.01a datasets. The empirical results indicate that DRSLSTM demonstrates superior average performance relative to all comparative algorithms, thereby confirming its enhanced efficacy and robustness across diverse experimental conditions. These empirical results unequivocally indicate that DRSLSTM achieves a marginally but consistently superior average performance compared to all six competing methods.

Table 4. Average Modulation Recognition Rates of 7 Methods in two datasets.

Dataset	CLDNN	Resnet	LSTM	CNN	Transformer	TLDNN	DRSLSTM
RML2016.01a	56.07%	61.33%	59.12%	57.14%	60.54%	62.82%	62.91%
RML2018.01a	84.63%	84.20%	77.87%	84.18%	83.31%	85.23%	86.30%

Note: Bold formatting indicates the best results.

4.4. The Ablation Study of DRSLSTM

The ablation study to provide a more detailed analysis of the contributions of the RSU and LSTM modules. The quantitative results of our ablation experiments are summarized in Table 5.

Table 5. Results of the Ablation Study on Model Components.

Model Configuration	Average Accuracy (%)
Full Model (RSU + LSTM)	86.30 %
Without LSTM Module	85.82%
Without RSU Module	85.09%

The results demonstrate that both modules contribute positively to the overall performance. The removal of the LSTM module leads to a performance drop of 0.48%, indicating its role in capturing temporal dependencies and sequential patterns. Conversely, excluding the RSU module results in a more significant reduction of 1.21%, suggesting its critical function in denoise. The RSU module acts as a feature pre-processor, providing cleaner and more salient features for the subsequent LSTM module. This allows the LSTM to more effectively model the temporal relationships based on high-quality input, rather than being distracted by noise. The results confirm that both components contribute uniquely and jointly to the model's robustness and accuracy.

5. Conclusions

This paper proposes a DRSLSTM network model suitable for recognizing signal modulation modes under low SNR conditions. The model employs a deep residual shrinkage module to denoise the original I/Q signals while extracting spatial features. Subsequently,

an LSTM network is utilized to capture long-short-term dependencies in the denoised features, thereby processing temporal features and recognizing modulation modes. Comparative experiments with six other network models demonstrate that the proposed DRSLSTM network model achieves a higher average recognition rate than the other models, and its performance is particularly superior to the six neural network models when the SNR is below 0 dB.

Certainly, while the proposed method has been validated on low-order modulation schemes, it encounters significant challenges when applied to high-order modulation schemes, such as 64QAM and OFDM. These challenges arise primarily from their heightened sensitivity to noise and inter-symbol interference, as well as the considerable complexity involved in feature extraction. Future work will focus on addressing these limitations by investigating hybrid approaches that integrate data-driven models with semantic attribute knowledge. This will involve leveraging higher-order cumulants and fractal dimensions to enhance robustness in noisy environments.

Author Contributions: Methodology, P.T. and D.C.; software, D.C.; validation, S.Z.; writing—original draft preparation, D.C.; writing—review and editing, K.Z. and Y.S.; visualization, S.Z.; investigation, supervision, and funding acquisition, P.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the following projects: Xiangjiang Laboratory (Grant Nos. 23XJ02006, 24XJJCYJ01001, 22XJ01002 and 23XJ01004), the Natural Science Foundation of Hunan Province (Grant Nos. 2025JJ50337 and 2023JJ50017), and the Research Foundation of Education Bureau of Hunan Province (Grant No. 23A0470).

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zheng, Q.; Tian, X.; Yu, L.; Elhanashi, A.; Saponara, S. Recent Advances in Automatic Modulation Classification Technology: Methods, Results, and Prospects. *Int. J. Intell. Syst.* **2025**, *2025*, 4067323. [[CrossRef](#)]
2. Huang, S.; He, J.; Yang, Z.; Chen, Y.; Chang, S.; Zhang, Y.; Feng, Z. Generalized Automatic Modulation Classification for OFDM Systems Under Unseen Synthetic Channels. *IEEE Trans. Wirel. Commun.* **2024**, *23*, 11931–11941. [[CrossRef](#)]
3. Xing, H.; Zhang, X.; Chang, S.; Ren, J.; Zhang, Z.; Xu, J.; Cui, S. Joint Signal Detection and Automatic Modulation Classification via Deep Learning. *IEEE Trans. Wirel. Commun.* **2024**, *23*, 17129–17142. [[CrossRef](#)]
4. Krzyston, J.; Bhattacharjea, R.; Stark, A. Complex-Valued Convolutions for Modulation Recognition Using Deep Learning. In Proceedings of the 2020 IEEE International Conference on Communications Workshops (ICC Workshops), Dublin, Ireland, 7–11 June 2020; pp. 1–6.
5. Li, L.; Dong, Z.; Zhu, Z.; Jiang, Q. Deep-Learning Hopping Capture Model for Automatic Modulation Classification of Wireless Communication Signals. *IEEE Trans. Aerosp. Electron. Syst.* **2023**, *59*, 772–783. [[CrossRef](#)]
6. Jiang, W.; Chen, J.; Jiang, Y.; Xu, Y.; Wang, Y.; Tan, L.; Liang, G. A New Time-Aware Collaborative Filtering Intelligent Recommendation System. *Comput. Mater. Contin.* **2019**, *61*, 849–859. [[CrossRef](#)]
7. Jiang, L.; Yu, S.; Wang, X.; Wang, C.; Wang, T. A New Source-Filter Model Audio Bandwidth Extension Using High Frequency Perception Feature for IoT Communications. *Concurr. Comput. Pract. Exp.* **2020**, *32*, e4638. [[CrossRef](#)]
8. Hazar, M.A.; Odabasioglu, N.; Ensari, T.; Kavurucu, Y.; Sayan, O.F. Performance Analysis and Improvement of Machine Learning Algorithms for Automatic Modulation Recognition over Rayleigh Fading Channels. *Neural Comput. Appl.* **2018**, *29*, 351–360. [[CrossRef](#)]
9. Han, L.; Gao, F.; Li, Z.; Dobre, O.A. Low Complexity Automatic Modulation Classification Based on Order-Statistics. *IEEE Trans. Wirel. Commun.* **2017**, *16*, 400–411. [[CrossRef](#)]
10. Xie, L.; Wan, Q. Cyclic Feature-Based Modulation Recognition Using Compressive Sensing. *IEEE Wirel. Commun. Lett.* **2017**, *6*, 402–405. [[CrossRef](#)]
11. Zhou, X.; Liang, W.; Yan, K.; Li, W.; Wang, K.I.K.; Ma, J.; Jin, Q. Edge-Enabled Two-Stage Scheduling Based on Deep Reinforcement Learning for Internet of Everything. *IEEE Internet Things J.* **2023**, *10*, 3295–3304. [[CrossRef](#)]

12. Mo, C.; Sun, W. Point-by-Point Feature Extraction of Artificial Intelligence Images Based on the Internet of Things. *Comput. Commun.* **2020**, *159*, 1–8. [\[CrossRef\]](#)
13. Ho, K.; Prokopiw, W.; Chan, Y. Modulation Identification of Digital Signals by the Wavelet Transform. *IEE Proc.—Radar Sonar Navig.* **2000**, *147*, 169–176. [\[CrossRef\]](#)
14. Walenczykowska, M.; Kawalec, A.; Krenc, K. An Application of Analytic Wavelet Transform and Convolutional Neural Network for Radar Intrapulse Modulation Recognition. *Sensors* **2023**, *23*, 1986. [\[CrossRef\]](#)
15. Zhou, X.; Xu, X.; Liang, W.; Zeng, Z.; Yan, Z. Deep-Learning-Enhanced Multitarget Detection for End-Edge-Cloud Surveillance in Smart IoT. *IEEE Internet Things J.* **2021**, *8*, 12588–12596. [\[CrossRef\]](#)
16. O'Shea, T.J.; West, N. Radio Machine Learning Dataset Generation with GNU Radio. Available online: <https://pubs.gnuradio.org/index.php/grcon/article/view/11/10> (accessed on 28 September 2025)
17. O'Shea, T.J.; Corgan, J.; Clancy, T.C. Convolutional Radio Modulation Recognition Networks. In Proceedings of the Engineering Applications of Neural Networks, Aberdeen, UK, 2–5 September 2016; pp. 213–226.
18. Erpek, T.; O'Shea, T.J.; Clancy, T.C. Learning a Physical Layer Scheme for the MIMO Interference Channel. In Proceedings of the 2018 IEEE International Conference on Communications (ICC), Kansas City, MO, USA, 20–24 May 2018; pp. 1–5.
19. O'Shea, T.J.; Roy, T.; West, N.; Hilburn, B.C. Demonstrating Deep Learning Based Communications Systems Over the Air In Practice. In Proceedings of the 2018 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN), Seoul, Republic of Korea, 22–25 October 2018; pp. 1–2.
20. Yang, G. Exponential Stability of Positive Recurrent Neural Networks with Multi-proportional Delays. *Neural Process. Lett.* **2019**, *49*, 67–78. [\[CrossRef\]](#)
21. Liu, X.; Wu, Z.; Tang, C. Modulation Recognition Algorithm Based on ResNet50 Multi-feature Fusion. In Proceedings of the 2021 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), Xi'an, China, 27–28 March 2021; pp. 677–680.
22. Qi, L.; Dou, W.; Hu, C.; Zhou, Y.; Yu, J. A Context-Aware Service Evaluation Approach over Big Data for Cloud Applications. *IEEE Trans. Cloud Comput.* **2020**, *8*, 338–348. [\[CrossRef\]](#)
23. Daldal, N.; Yildirim, Ö.; Polat, K. Deep Long Short-Term Memory Networks-Based Automatic Recognition of Six Different Digital Modulation Types under Varying Noise Conditions. *Neural Comput. Appl.* **2019**, *31*, 1967–1981. [\[CrossRef\]](#)
24. Kong, W.; Yang, Q.; Jiao, X.; Niu, Y.; Ji, G. A Transformer-based CTDNN Structure for Automatic Modulation Recognition. In Proceedings of the 2021 7th International Conference on Computer and Communications (ICCC), Chengdu, China, 10–13 December 2021; pp. 159–163.
25. Wang, Y.; Fang, S.; Fan, Y.; Wang, M.; Xu, Z.; Hou, S. A Complex-Valued Convolutional Fusion-Type Multi-Stream Spatiotemporal Network for Automatic Modulation Classification. *Sci. Rep.* **2024**, *14*, 22401. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Zhang, W.T.; Cui, D.; Lou, S.T. Training Images Generation for CNN Based Automatic Modulation Classification. *IEEE Access* **2021**, *9*, 62916–62925. [\[CrossRef\]](#)
27. Xue, W. Anti-Interference Performance of QPSK Modulation and Demodulation Technology in Mobile Communication with MATLAB Simulation. In Proceedings of the 2025 IEEE 5th International Conference on Electronic Technology, Communication and Information (ICETCI), Changchun, China, 23–25 May 2025; pp. 176–181.
28. Zhang, J.; Bhuiyan, M.Z.A.; Yang, X.; Wang, T.; Xu, X.; Hayajneh, T.; Khan, F. AntiConcealer: Reliable Detection of Adversary Concealed Behaviors in EdgeAI-Assisted IoT. *IEEE Internet Things J.* **2022**, *9*, 22184–22193. [\[CrossRef\]](#)
29. Morohashi, I.; Sekine, N. Generation and Detection of FM-CW Signals in All-Photonic THz Radar Systems. In Proceedings of the 2024 49th International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz), Perth, Australia, 1–6 September 2024; pp. 1–2.
30. Jiang, L.; Hu, R.; Wang, X.; Tu, W.; Zhang, M. Nonlinear Prediction with Deep Recurrent Neural Networks for Non-Blind Audio Bandwidth Extension. *China Commun.* **2018**, *15*, 72–85. [\[CrossRef\]](#)
31. Zhang, S.; Chen, A.; Guo, W.; Cui, Y.; Zhao, X.; Liu, L. Learning Deep Binaural Representations With Deep Convolutional Neural Networks for Spontaneous Speech Emotion Recognition. *IEEE Access* **2020**, *8*, 23496–23505. [\[CrossRef\]](#)
32. Jiang, J.; Wang, Z.; Zhao, H.; Qiu, S.; Li, J. Modulation Recognition Method of Satellite Communication Based on CLDNN Model. In Proceedings of the 2021 IEEE 30th International Symposium on Industrial Electronics (ISIE), Kyoto, Japan, 20–23 June 2021; pp. 1–6.
33. Qu, Y.; Lu, Z.; Zeng, R.; Wang, J.; Wang, J. Enhancing Automatic Modulation Recognition Through Robust Global Feature Extraction. *IEEE Trans. Veh. Technol.* **2025**, *74*, 4192–4207. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.