TriFusion-Lite: A Temporal-Frequency-Phase Fusion Lightweight Network for Modulation Recognition

Rujiao CHENG [1,*], *Juan SU* [1], *Min HUANG* [2]

[1] School of Electrical and Electronic Engineering, Chengdu Technological University, Chengdu 611700, China
[2] School of Intelligent Technology, TianFu College of Southwestern University of Finance and Economics, MianYang 621050, China

*crjiao1@cdtu.edu.cn

Submitted January 16, 2026 / Accepted March 27, 2026 / Online first May 26, 2026

Abstract. *Automatic Modulation Recognition (AMR) is an essential technology for modern wireless communication systems. However, existing deep learning models are often computationally complex and frequently overlook the phase relationships within in-phase/quadrature (I/Q) signals, thereby hindering their deployment on resource-constrained devices. This paper introduces TriFusion-Lite, a lightweight, multi-stream deep learning architecture proposed for optimizing both the efficiency and precision in AMR. The proposed framework initiates with a tailored preprocessing pipeline that compresses the input signal while enriching its representation with robust statistical features. A novel four-stream parallel network then processes this enhanced signal: 1) a complex-valued convolutional stream to preserve phase integrity; 2) two parallel 1D convolutional streams for independent I/Q channel analysis; and 3) a Short-Time Fourier Transform (STFT) stream to capture spectral characteristics. A hierarchical fusion mechanism then progressively integrates these multi-domain features for final classification. Comprehensive evaluations on benchmark datasets demonstrate the effectiveness and competitive performance of our approach. Our experiments confirm the effectiveness of compression, showcasing its performance characteristics across various compression levels. Furthermore, the proposed method achieves competitive results compared to state-of-the-art approaches, striking an effective balance between performance and computational efficiency, thereby presenting a promising approach for AMR applications on edge devices. The source code of the proposed framework is publicly available at https://github.com/sansi34jun/TriFusion-Lite.*

Keywords

Modulation recognition, I/Q signals, lightweight neural network, multi-stream fusion

1. Introduction

AMR is a foundational technology in modern wireless communication systems. It identifies how a received signal is modulated without needing prior information. This ability is essential for dynamic spectrum access, interference identification, adaptive communications, and intelligent resource management in 5G/6G networks [1]. Traditionally, AMR methods are typically categorized into two main types: likelihood-based (LB) and feature-based (FB) approaches. LB methods, such as the average likelihood-ratio test, are considered theoretically optimal under known channel conditions. However, they are highly dependent on accurate channel information, which is frequently unavailable in real-world, non-cooperative communication scenarios [2]. FB methods address this limitation by extracting statistical properties from the signal. For instance, Higher-Order Cumulants (HOCs) have been employed due to their insensitivity to Gaussian noise, which aids in distinguishing different signal shapes [3]. Other early approaches utilized cyclostationarity to analyze modulated signals, as shown in reference [4], a method aimed at identifying the periodic patterns present in the signal. However, these FB methods rely on carefully designed "golden features," which demands expert knowledge. This limitation has motivated the development of more flexible, data-driven deep learning solutions.

The advent of deep learning, with its powerful hierarchical feature extraction capabilities, has revolutionized traditional machine learning and achieved state-of-the-art performance in diverse domains such as speech recognition [5] and computer vision [6]. This paradigm shift has naturally extended to the AMR field, enabling end-to-end feature learning directly from raw *I/Q* data. Pioneering work by O'Shea et al. [7] demonstrated the potential of Convolutional Neural Network (CNN) based models to outperform traditional methods on standard datasets. Inspired by this breakthrough, the research community rapidly advanced the field by exploring more sophisticated architectures. Furthermore, other work has employed Long Short-Term Memory (LSTM) networks to learn directly from the signal's amplitude and phase

DOI: 10.13164/re.2026.0304

features, thus bypassing the complex process of expert feature extraction [8]. Furthermore, deeper network structures, such as Residual Networks (ResNet) [9] and Convolutional, Long Short-Term Memory, Deep Neural Networks (CLDNN) [10], have pushed the boundaries of classification accuracy. While these advanced models have achieved notable success, their performance gains are often accompanied by increased computational demands.

Despite these achievements, a closer examination reveals persistent challenges that hinder the practical deployment of many existing DL-based AMR models. First, the issue of phase information loss remains prevalent. Most architectures treat the complex-valued *I/Q* signal as two independent real-valued channels, a simplification that discards phase relationships, which are essential for distinguishing between modulations such as phase-shift keying. While some efforts have incorporated complex-valued operations, including our own prior work on the CVCNN-LSTM architecture [11] and other complex-valued networks [12], [13], we observed that these models often introduce significant computational complexity or require specialized libraries, limiting their broad applicability and motivating our current lightweight design. Second, the fusion of features from different domains is often suboptimal. Many approaches either rely solely on time-domain features [14] or employ simple concatenation of temporal and spectral representations, a static fusion strategy that overlooks the dynamic interplay between domains [15]. While more advanced techniques incorporating attention mechanisms [16] or spatiotemporal learning frameworks [17] have shown promise, they often lead to a significant increase in model complexity and computational overhead. Recent studies continue to explore more sophisticated fusion methods, such as cross-modal attention and transformer-based architectures [18], [19], yet achieving an effective and efficient fusion remains a key challenge. Recent studies have explored advanced techniques to address these issues. For example, graph attention networks have shown promise in low Signal-to-Noise Ratio (SNR) scenarios [20], while dual-input architectures combining *I/Q* and amplitude-phase data with knowledge distillation have achieved strong performance across multiple datasets [21]. Additionally, systematic comparisons of CNN, GRU, and hybrid models under various RF impairments highlight the importance of dataset diversity and model selection [22]. However, these methods often incur high computational cost or lack effective multi-domain fusion, motivating our lightweight TriFusion-Lite design that jointly leverages temporal, frequency, and phase information.

A significant challenge for the practical application of AMR is the computational efficiency of deep learning models. Many state-of-the-art architectures, while achieving high accuracy, are parameter-heavy and computationally intensive [23], rendering them unsuitable for deployment on resource-constrained edge devices such as Internet of Things (IoT) sensors or unmanned aerial vehicles. This "performance-efficiency gap" has motivated a growing body of research focused on developing lightweight architectures that maintain robust performance [24]. Methodologies for model compression, including knowledge distillation, network pruning, and quantization, have recently been explored to reduce the footprint of AMR models [25]. However, these techniques often compress a pre-existing large model, whereas designing an intrinsically efficient architecture that synergistically integrates multi-domain features from the ground up remains a more fundamental challenge. To systematically address these multifaceted challenges, we propose TriFusion-Lite, an innovative framework designed from the ground up for efficient and accurate AMR on edge devices. The name TriFusion-Lite reflects its core principles: temporal-frequency-phase fused feature processing and a lightweight network architecture. Our work introduces a holistic solution that spans from intelligent signal preprocessing to a novel multi-stream fusion architecture. The main contributions of this paper are summarized as follows:

1. We propose a tailored preprocessing pipeline that compresses the input signal while extracting robust statistical features, including HOCs.

2. We introduce a novel, efficiency-centric parallel network with four distinct streams: a complex-valued stream for phase integrity, two 1D convolutional streams for *I/Q* amplitude dynamics, and an STFT stream for spectral patterns. These multi-domain features are integrated via hierarchical fusion, learning comprehensive and discriminative representations, utilizing techniques like depthwise separable convolutions.

3. Our approach is rigorously validated on RML2016.10a, RML2016.10b, and RML2018.01a datasets. The model achieves state-of-the-art or highly competitive accuracy with significantly fewer parameters and faster training, balancing performance and computational efficiency.

The remainder of this paper is organized as follows. Section 2 describes the proposed TriFusion-Lite framework in detail. Section 3 presents the experimental setup, datasets, results, and discussion. Section 4 concludes the paper and outlines future work.

2. Proposed Method

The main contributions of TriFusion-Lite are: (i) a novel temporal-frequency-phase fusion mechanism, (ii) a lightweight architecture suitable for edge devices, and (iii) state-of-the-art accuracy with significantly lower computational cost. The TriFusion-Lite architecture is based on a multi-domain design for lightweight AMR. We hypothesize that robust classification, especially in low SNRs regimes, necessitates the synergistic integration of features from the temporal, spectral, and complex-phase domains. To this end, our framework is structured into two primary stages: a tailored data preprocessing pipeline for feature enhancement, and a novel four-stream deep learning architecture for hierar-

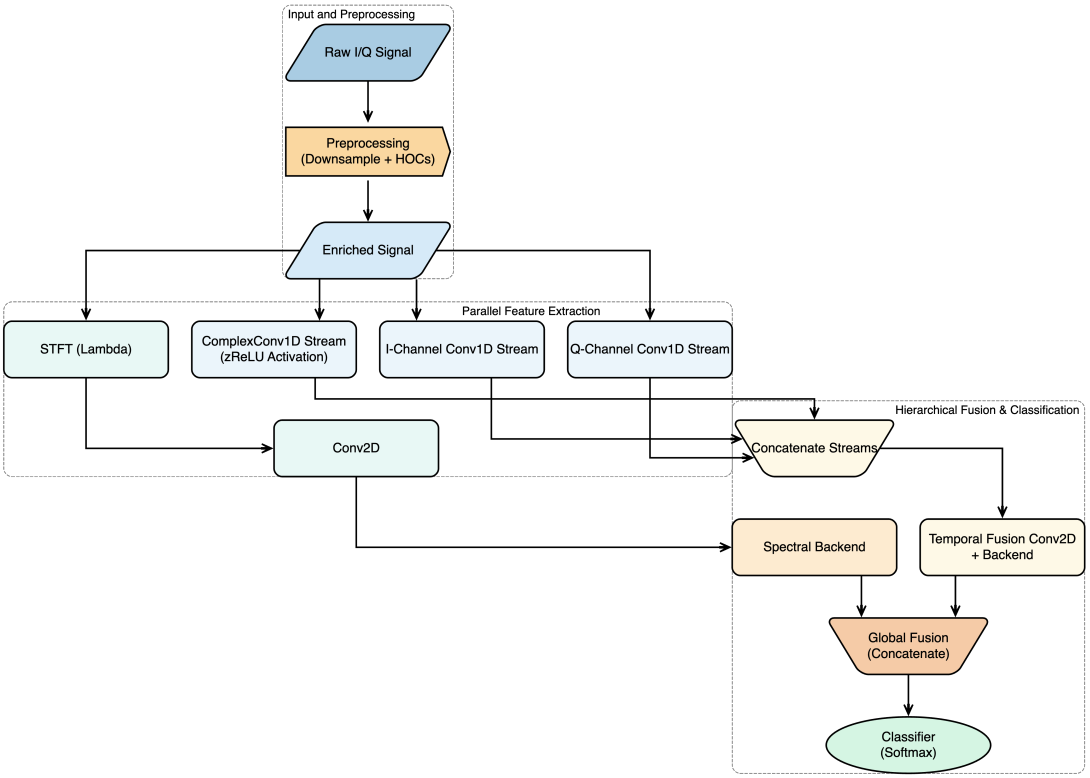


Fig. 1. Overall block diagram of the proposed TriFusion-Lite framework.

Layer (Group)	Output shape	Params	Core configuration details
Input	$(2, 69)$	0	Enriched *I/Q* signals after preprocessing ($CF = 2$).
Complex stream	$(63, 60)$	480	Kernel=7, Filters=60, zReLU activation.
I/Q streams	$2 \times (63, 60)$	960	Parallel Conv1D, Kernel=7, Filters=60.
STFT branch	$(7, 17, 64)$	1,216	STFT + Conv2D (Kernel=3×3, Filters=64).
Temporal fusion	$(57, 30)$	49,580	Fuses *I/Q*/Complex streams via Conv2D, then reduces.
Global fusion	(192)	55,680	Fuses temporal path and STFT path.
Classifier	(11)	15,719	Hidden Dense (77 units) + Softmax output.
Total		**123,635**	

Tab. 1. Detailed configuration of the TriFusion-Lite architecture.

chical feature extraction and fusion. A high-level overview of the complete architecture is presented in Fig. 1. The following subsections will detail each of these core components, with the final layer-wise configuration summarized in Tab. 1.

The TriFusion-Lite architecture is motivated by the observation that modulation information resides in multiple domains: temporal amplitude variations, spectral energy distribution, and phase relationships. To capture these complementary features without incurring high computational cost, we design a four-stream parallel structure. The complex-valued stream preserves phase integrity through complex convolutions and zReLU activation. Two independent 1D convolutional streams extract temporal dynamics from the I and Q channels separately. The STFT-based spectral stream captures time-frequency patterns via 2D convolutions. These streams operate on a compressed and feature-enhanced input produced by our preprocessing pipeline. A two-stage hierarchical fusion mechanism first integrates the three temporal streams, then combines the result with spectral features, enabling the model to learn cross-domain correlations efficiently. The entire architecture employs lightweight components such as depthwise separable convolutions to maintain low parameter count.

2.1 Data Preprocessing and Feature Enhancement

To reconcile the conflicting demands of computational efficiency and feature preservation, we introduce a tailored preprocessing pipeline.This pipeline is designed to transform a raw *I/Q* signal, for instance, a sample $\mathbf{Z} \in \mathbb{R}^{2\times 128}$ from the RML2016.10a dataset [7], into an enriched and compressed representation $\mathbf{F} \in \mathbb{R}^{2\times M}$ that is optimized for our lightweight network. The length of the compressed signal, M, is determined by a compression factor (CF), which will be investigated in our experiments.

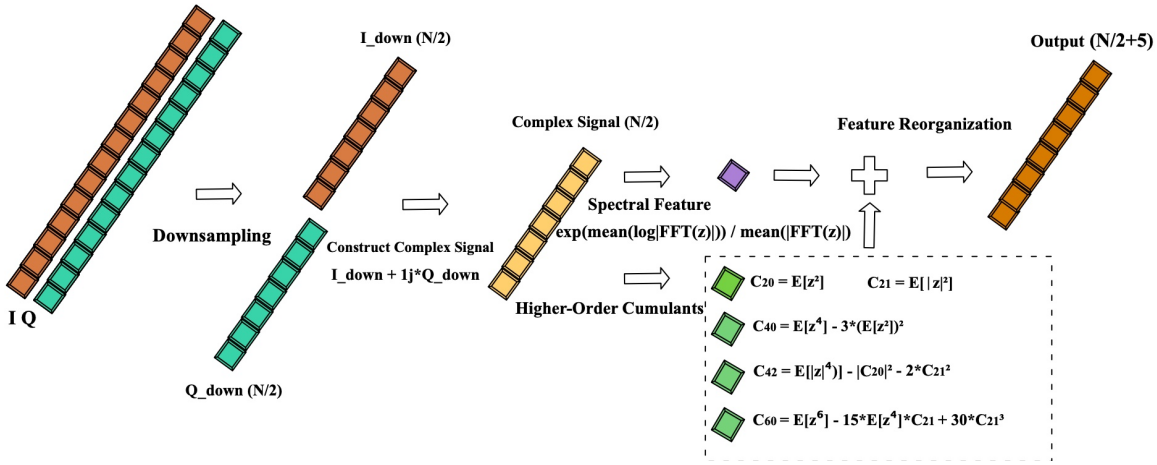
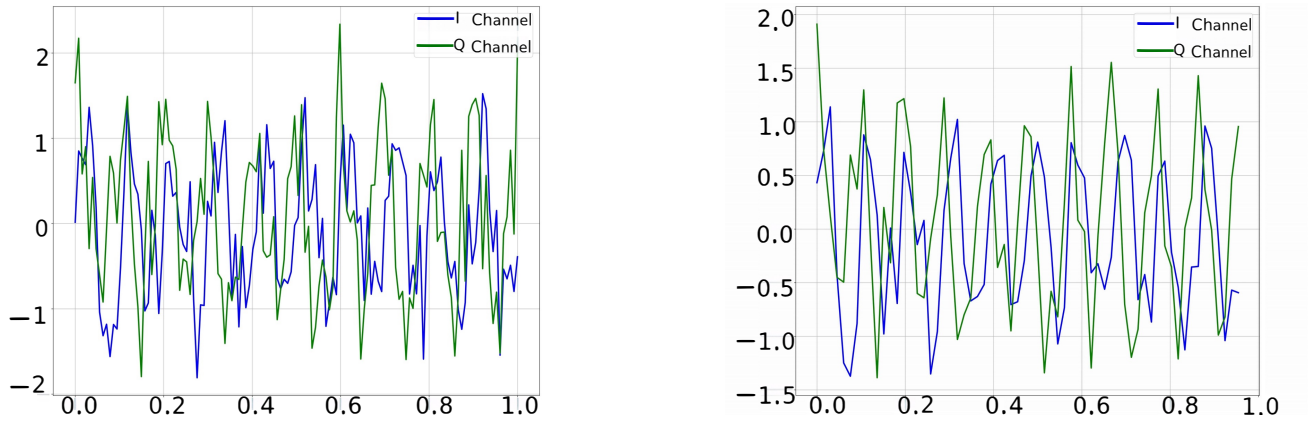


Fig. 2. Data preprocessing pipeline illustrating downsampling, statistical feature extraction, and concatenation.



(a) Original *I/Q* signal ($N = 128$) (b) Compressed ($M = 69$, $CF = 2$)

Fig. 3. Visualization of an *I/Q* signal from the RML2016.10a dataset before and after applying the proposed preprocessing.

As illustrated in Fig. 2, the first module in the overall framework (Fig. 1) is the preprocessing block. This block performs workflow, visually detailed in Fig. 2, begins with signal compression. Raw I and Q components, each with $N = 128$ samples, are downsampled by a factor of CF. Instead of simple decimation, we apply a 1D convolution with a moving-average kernel of size CF, acting as a low-pass anti-aliasing filter to preserve crucial low-frequency envelopes. For $CF = 2$, the sequence length reduces to $N' = N/CF = 128/2 = 64$ samples.

Following compression, we augment the downsampled signals (I_{down} and Q_{down}) with high-level statistical features to compensate for temporal resolution reduction. To analyze the signal more reliably, we extract five selected, stable statistical features that can resist noise interference:

1. Spectral Flatness (S_{F}): S_{F}, measuring the flatness of the signal's power spectrum.

2. HOCs: $C_{20}, C_{40}, C_{42}, C_{60}$, powerful for characterizing non-linearities and symmetries of a signal's probability distribution [26].

These statistical descriptors are then selectively concatenated to form enriched feature vectors for the I and Q channels. The augmented I-channel sequence, $\mathbf{I}_{\text{enhanced}}$, and Q-channel sequence, $\mathbf{Q}_{\text{enhanced}}$, are formulated as:

$$\mathbf{I}_{\text{enhanced}} = [I_{\text{down}}, C_{20.\text{real}}, C_{40.\text{real}}, C_{42.\text{real}}, C_{60.\text{real}}, S_{\text{F}}], \quad (1)$$

$$\mathbf{Q}_{\text{enhanced}} = [Q_{\text{down}}, C_{20.\text{imag}}, C_{40.\text{imag}}, C_{42.\text{imag}}, 0, 0]. \quad (2)$$

Here, I_{down} and Q_{down} represent the downsampled I and Q sequences, respectively. The dimensions are maintained by padding the Q-channel with two zeros for consistency. This results in an enriched feature vector $\mathbf{Z}_{\text{preprocessed}}$ for each I and Q channel, both with a total length of $M = N' + 5$.

For our primary configuration with $CF = 2$, this yields an individual I or Q sequence length of $M = 64 + 5 = 69$, as visualized in Fig. 3. Similarly, for $CF = 3$ and $CF = 4$, the resulting lengths would be $\lfloor 128/3 \rfloor + 5 = 42 + 5 = 47$ and $\lfloor 128/4 \rfloor + 5 = 32 + 5 = 37$, respectively. This entire preprocessing stage is crucial for distilling high-value information into a compact format amenable to a lightweight network. The impact of varying this compression factor on model performance and efficiency will be evaluated in Sec. 3.2.

2.2 Four-Stream Parallel Feature Extraction

The cornerstone of our architecture is a four-stream parallel network, where each stream is a specialized feature extractor dedicated to a distinct signal domain. This design is motivated by the principle of feature disentanglement, allowing the model to learn complementary representations simultaneously. A similar concept of leveraging parallel multiscale structures to enhance feature diversity while maintaining efficiency has also been validated in recent work [27].

Complex-valued convolutional stream: This stream is engineered to holistically preserve the interplay between signal magnitude and phase. A custom `ComplexConv1D` layer efficiently implements complex convolution. Let a complex kernel be $\mathbf{W} = \mathbf{W}_r + \mathrm{j}\mathbf{W}_i$ and a complex input be $\mathbf{X} = \mathbf{I} + \mathrm{j}\mathbf{Q}$. The output is:

$$\mathbf{Y} = (\mathbf{I} * \mathbf{W}_r - \mathbf{Q} * \mathbf{W}_i) + \mathrm{j}(\mathbf{Q} * \mathbf{W}_r + \mathbf{I} * \mathbf{W}_i). \quad (3)$$

A key component of this stream is the subsequent application of the quadrant-limited zReLU activation function, a concept we introduced and validated in our prior work [11]. The zReLU function is defined as:

$$\mathrm{zReLU}(z) = \begin{cases} z, & \text{if } \mathrm{Re}(z) \geq 0, \mathrm{Im}(z) \geq 0, \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

As visualized in the complex plane, zReLU imposes a phase-selective non-linearity. This targeted activation breaks the inherent symmetry of noise distributions, which are often centered at the origin, effectively functioning as a non-linear noise filter that enhances model robustness.

Parallel *I/Q* temporal streams: Complementing the complex stream, two parallel streams apply standard `Conv1D` layers to the I and Q channels independently. This is motivated by the need to explicitly learn the marginal properties and amplitude-specific dynamics crucial for non-constant modulus modulations like QAM and ASK.

STFT spectral stream:A fourth stream analyzes the signal's spectral morphology via the STFT. This maps the 1D signal into a 2D time-frequency representation, allowing a `Conv2D` layer to operate as a learned spectral shape detector, identifying features orthogonal to the time-domain streams.

2.3 Hierarchical Feature Fusion and Classification

As depicted in Fig. 1, the model employs a two-stage hierarchical mechanism to fuse the multi-domain features.

In the first stage, the outputs of the three temporal streams (Complex, I-channel, and Q-channel) are integrated. A convolutional layer is applied across these stacked streams to learn local, cross-stream correlations, effectively fusing phase and amplitude dynamics into a unified temporal representation.

In the second stage, this fused temporal representation and the features from the STFT stream are processed through their respective network backends. Global Average Pooling then distills each path into a compact summary vector. These two vectors, representing the temporal-phase ($\mathbf{v}_{\text{temp}}$) and spectral ($\mathbf{v}_{\text{spec}}$) characteristics, are concatenated to form a comprehensive final representation:

$$\mathbf{v}_{\text{final}} = [\mathbf{v}_{\text{temp}} \| \mathbf{v}_{\text{spec}}]. \quad (5)$$

This final vector is then passed to a classifier, which uses a Softmax function to make the final modulation decision. This progressive fusion strategy creates a highly discriminative representation for robust classification.

2.4 Implementation Details

To provide a concrete instantiation of our proposed framework, the complete architectural details are summarized in Tab. 1. The table specifies the layer types, output shapes, parameter counts, and core configurations for the model as applied to the RadioML2016.10a dataset. It is important to note that the network's input shape is $(2, 69)$, which is the result of our data preprocessing pipeline applied to the original $(2, 128)$ signals.

3. Experimental Evaluation

To validate the performance and efficiency of the proposed method, we conduct a series of experiments. Our evaluation is structured into two main parts: first, an analysis of the impact of the input signal's compression factor on model performance, detailed in Sec. 3.2. Second, a thorough benchmarking of our model against several state-of-the-art methods across three distinct datasets, presented in Sec. 3.4. The experiments are designed to assess three key metrics: classification accuracy (effectiveness), model size (parameter count), and training speed (computational efficiency).

3.1 Dataset and Preprocessing

Our experimental framework utilizes three benchmark datasets for modulation recognition, each presenting distinct characteristics: RML2016.10a, RML2016.10b [7], and RML2018.01a [9]. RML2016.10a offers simulated signals (2×128) across 11 modulation types, with realistic channel impairments including additive white Gaussian noise (AWGN), multipath fading, carrier frequency offset, and sampling rate offset. Its expanded counterpart, RML2016.10b, retains 10 modulation categories but scales the sample size to 1.2 million for robustness assessment. Conversely, RML2018.01a provides laboratory-captured signals (2×1024) with 24 modulation types and precise channel control, facilitating high-fidelity evaluation of temporal modeling.

To balance computational feasibility on the Apple M3 Max SoC (12-core CPU, 40-core GPU, 128GB unified memory) with dataset representativeness, we employ the complete datasets for RML2016.10a (220K samples) and RML2016.10b (1.2M samples). For the larger RML2018.01a, a stratified sampling approach retaining 10% of the data are applied, meticulously preserving the original class distribution and noise-level characteristics. This strategy enables comprehensive performance assessment across varying modulation complexities, noise conditions, and sequence lengths.

All datasets are subjected to identical preprocessing: a 60%/20%/20% stratified split into training, validation, and test sets, ensuring class balance. Models are trained using categorical cross-entropy loss and the Adam optimizer (initial learning rate = 0.001), with early stopping implemented (monitoring validation loss over 10 epochs patience). Experiments are conducted on macOS Ventura 13.4, with tensor computations accelerated via the Apple Metal API optimized for M-series silicon. The models are implemented using PyTorch 2.0.1 with the Metal Performance Shaders backend.

To evaluate the performance of our proposed method, we use classification accuracy as the primary metric. In this work, classification accuracy is defined as the ratio of correctly predicted samples to the total number of samples in the test set. This metric provides a straightforward measure of the model's overall recognition performance and is widely adopted in the AMR literature [7], [28]

3.2 Compression Factor Analysis

We systematically evaluate compression factors $CF = \{2, 3, 4\}$ against the uncompressed baseline using the RadioML2016.10a dataset. The original 128-sample *I/Q* sequences are compressed to lengths $N = \lfloor 128/CF{+}5 \rfloor$ through adaptive downsampling. Table 2 illustrates the relationship between accuracy, inference time, and the degree of model compression.

The results demonstrate that a compression factor of $CF = 2$ achieves a favorable balance between performance and efficiency. On the RML2016.10a dataset, this configuration reduces the sequence length to 69 samples, yet it achieves an average accuracy of 62.63%, outperforming the uncompressed baseline's accuracy of 62.10%. This slight increase in accuracy is attributed to the downsampling kernel acting as a low-pass filter, which smooths out high-frequency noise while preserving essential modulation features. The $CF = 2$ setting significantly improves computational efficiency, reducing the average epoch duration from 6.4 s to 3.7 s.

CF	Seq. len.	Epoch duration [s]	Avg. accuracy [%]
Baseline	128	6.4	62.10
2	69	3.7	62.63
3	47	3.2	61.28
4	37	3.2	60.05

Tab. 2. Performance trade-off analysis for different *CF* on RML2016.10a.

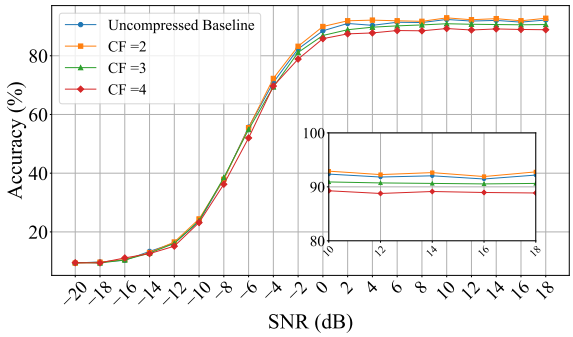


Fig. 4. Modulation recognition accuracy vs. SNR under different *CF*.

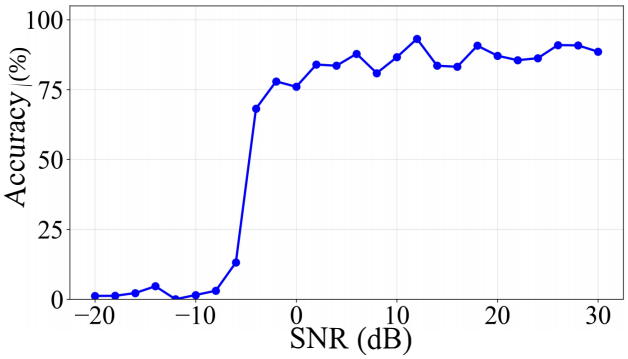


Fig. 5. Recognition accuracy of 256-QAM modulation on the RML2018.01a dataset.

As shown in Fig. 4, the $CF = 2$ configuration consistently achieves higher accuracy across most SNRs from –20 dB to 20 dB, significantly outperforming both $CF = 3$ and $CF = 4$. At low Signal-to-Noise Ratio (SNR) regions, the accuracy of all configurations is relatively low. However, as SNR increases, accuracy significantly improves across the board. Notably, when SNR exceeds –4 dB, the $CF = 2$ configuration demonstrates a clear advantage in accuracy, ultimately achieving near-perfect accuracy in high SNR regions. This indicates that under favorable SNR conditions, the $CF = 2$ configuration delivers superior modulation recognition performance. While some challenges remain at low SNRs, the prominent advantage of $CF = 2$ at high SNRs highlights its significant value in practical applications.

To assess the capability of TriFusion-Lite on higher-order modulations, we evaluate its performance on the 256-QAM signals contained in the RML2018.01a dataset. Figure 5 shows the recognition accuracy of 256-QAM across different SNR levels. The model achieves a reasonable accuracy at high SNRs, demonstrating its potential for handling complex modulations. However, the performance is not yet saturated; further research will focus on improving the recognition of high-order QAM variants.

3.3 Model Component Ablation

To systematically evaluate the contribution of each stream in TriFusion-Lite, we conduct ablation studies by removing three key components: the complex-valued stream (TriFusion-Lite-A), the STFT spectral stream (TriFusion-

Configuration	Avg. accuracy [%]	Drop [%]
TriFusion-Lite (Full)	62.6	-
TriFusion-Lite-A	57.8	4.8
TriFusion-Lite-B	60.3	2.3
TriFusion-Lite-C	57.3	5.3

Tab. 3. Performance comparison of the TriFusion-Lite with its varieties.

Lite-B), and one of the parallel *I/Q* temporal streams (TriFusion-Lite-C). Table 3 quantifies the overall performance metrics on the RML2016.10a dataset, demonstrating the impact of each component on classification accuracy. The results show that each component contributes positively to the overall performance, with the complex-valued stream having the most significant impact. The full TriFusion-Lite achieves the highest accuracy, confirming the effectiveness of its multi-stream collaborative design. It should be noted that while the average accuracy of TriFusion-Lite is approximately 62.6%, this performance is highly competitive within the specialized scope of lightweight architectures. Specifically, as evidenced by the high-SNR regions in Fig. 6, the model achieves near-optimal performance, with the lower average primarily stemming from the inherent challenges of deep-noise regimes (SNR < –10 dB). This metric reflects a cumulative average across all tested SNR levels, including those with extremely poor signal quality. As further demonstrated in the comparisons with other state-of-the-art methods (e.g., in Tab. 4), our model achieves a superior balance between recognition precision and computational efficiency, proving its leading position in resource-constrained AMR tasks.

3.4 State-of-the-Art Comparison

To validate the effectiveness of TriFusion-Lite, we conduct comparative experiments with three state-of-the-art AMR methods: CVCNN-LSTM [11], PETCGDNN [29], MCLDNN [17], and TLDNN [30]. The evaluation is performed on three standard benchmark datasets: RML2016.10a, RML2016.10b, and RML2018.01a. For fair comparison, all baseline methods are reproduced using the same training/validation/test splits and evaluated under identical hardware conditions. Hyperparameters for each baseline are kept as reported in their original papers. The key performance metrics—parameter count, training efficiency (average epoch duration), and classification accuracy—are summarized in Tab. 4.

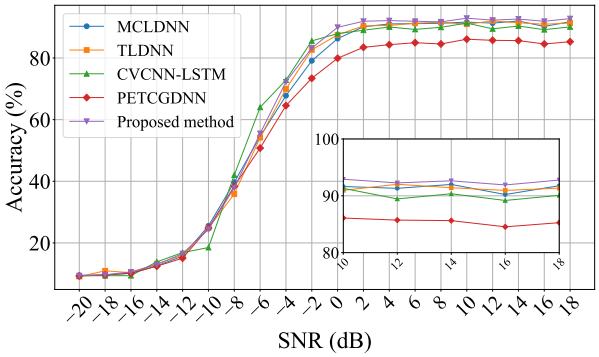
As shown in Tab. 4, our model demonstrates a favorable balance between performance and computational cost. On the RML2016.10a dataset, our method achieves the highest average accuracy of 62.6%, surpassing all benchmark models. Notably, compared to MCLDNN, our model uses 69% fewer parameters. On the RML2016.10b dataset, our model's average accuracy of 64.4% matches the performance of the best-performing baseline, MCLDNN. This state-of-the-art accuracy is achieved with only one-third of the parameters required by MCLDNN, highlighting the exceptional parameter efficiency of our architecture.For the more complex RML2018.01a dataset, our model remains highly competitive, with an average accuracy of 57.0%. It significantly outperforms MCLDNN. While other models may show a slightly higher average accuracy, our model provides a more favorable trade-off between speed and accuracy. Furthermore, we evaluate the real-time feasibility of TriFusion-Lite. The model achieves an inference time of 2.3 ms per sample on a CPU (Apple M3). While this exceeds the 0.5–1 ms slot duration of 5G NR, it remains acceptable for many edge applications with less stringent latency requirements, such as IoT sensors and drone communications. With batch processing or deployment on lightweight hardware accelerators (e.g., FPGA, Edge TPU), the latency can be further reduced to sub-millisecond levels. These results demonstrate that TriFusion-Lite offers a practical trade-off between accuracy and real-time performance for real-world AMR tasks.

As shown in Fig. 6, our model consistently maintains a more competitive edge than other state-of-the-art methods across all three datasets and different SNRs.On the RML2016.10a dataset (Fig. 6(a)), our model performs exceptionally well in the mid-to-low SNR region from –8 dB to 10 dB, highlighting its robustness in noisy environments. For the larger RML2016.10b dataset (Fig. 6(b)), our model's average accuracy reaches 64.4%, matching the performance of the best-performing MCLDNN model.It particularly excels in high SNR regions above 10 dB, where it consistently outperforms all other methods, indicating excellent performance in clear channel conditions. On the more complex RML2018.01a dataset (Fig. 6(c)), the model demonstrates highly competitive performance, achieving high accuracy at high SNRs, similar to that observed with PETCGDNN. Collectively, these results suggest our approach offers a well-balanced and robust performance across a range of challenging conditions. As illustrated in Fig. 7, the TriFusion-Lite model exhibits rapid convergence during the initial 20 epochs, with both accuracy and loss curves stabilizing shortly thereafter. Notably, the validation accuracy consistently tracks or slightly exceeds the training accuracy, which validates the efficacy of our lightweight design and the inclusion of dropout/regularization in preventing overfitting. Despite the inherent volatility of *I/Q* signal data, the smooth convergence trend indicates that the multi-stream fusion architecture (temporal, frequency, and phase) provides a stable gradient flow, ensuring robust learning dynamics even with a limited parameter budget.

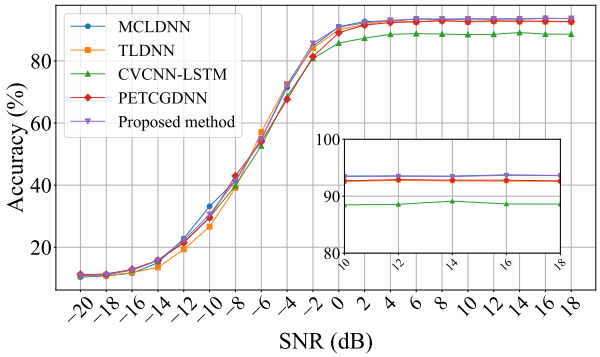
A qualitative analysis of the model's behavior is provided by the confusion matrices in Fig. 8. At extremely low SNRs (–20 dB and –10 dB), the model struggles, which is an expected behavior. However, as SNR improves to 0 dB (Fig. 8(c)), the model achieves 90.0% accuracy, with most confusion occurring between spectrally similar modulations (e.g., 16-QAM vs. 64-QAM). At high SNRs (≥ 10 dB), the confusion matrices (Fig. 8(d–e)) exhibit strong diagonal dominance with over 92% accuracy, indicating excellent discriminative capability. The overall matrix (Fig. 8(f)) confirms that the model robustly identifies most modulation types, establishing its effectiveness for practical AMR tasks.

Datasets	Method	Parameters	Avg. epoch duration [s]	Highest accuracy [%]	Avg. accuracy [%]
RML2016.10a	PETCGDNN	71,871	3.8	86.1	57.6
	TLDNN	259,723	8.0	92.0	61.8
	MCLDNN	405,175	13.0	92.0	61.5
	CVCNN-LSTM	325,898	9.3	91.3	61.9
	Proposed method	**123,635**	**3.7**	**92.9**	**62.6**
RML2016.10b	PETCGDNN	71,742	30.0	92.9	62.6
	TLDNN	259,690	40.4	93.0	63.5
	MCLDNN	405,046	64.4	**93.8**	**64.4**
	CVCNN-LSTM	325,820	50.9	89.1	61.3
	Proposed method	**123,557**	**26.9**	93.7	**64.4**
RML2018.01a	PETCGDNN	75,340	60.4	**92.1**	**57.3**
	TLDNN	280,760	29.4	85.9	53.2
	MCLDNN	406,852	132.2	90.0	55.7
	CVCNN-LSTM	330,368	102.3	91.2	56.6
	Proposed method	**124,649**	**60.0**	91.9	57.0

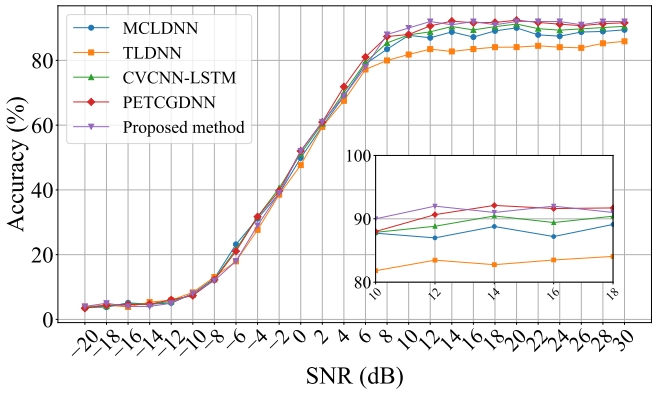
Tab. 4. Performance comparison with state-of-the-art methods on benchmark datasets.



(a) Performance on RML2016.10a



(b) Performance on RML2016.10b

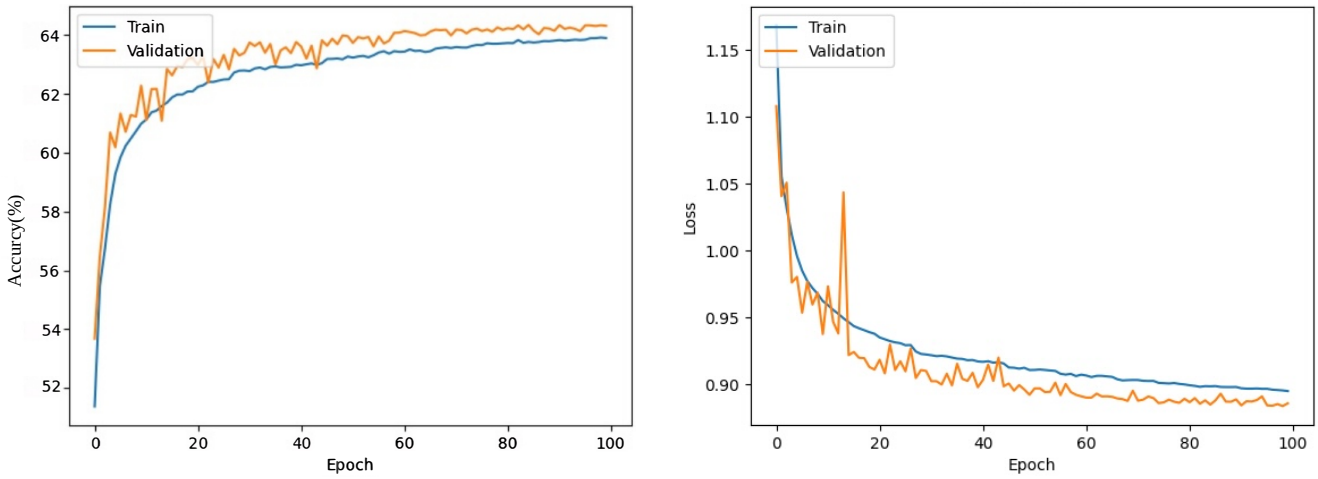


(c) Performance on RML2018.01a

Fig. 6. Accuracy comparison of TriFusion-Lite versus other methods across SNR levels.

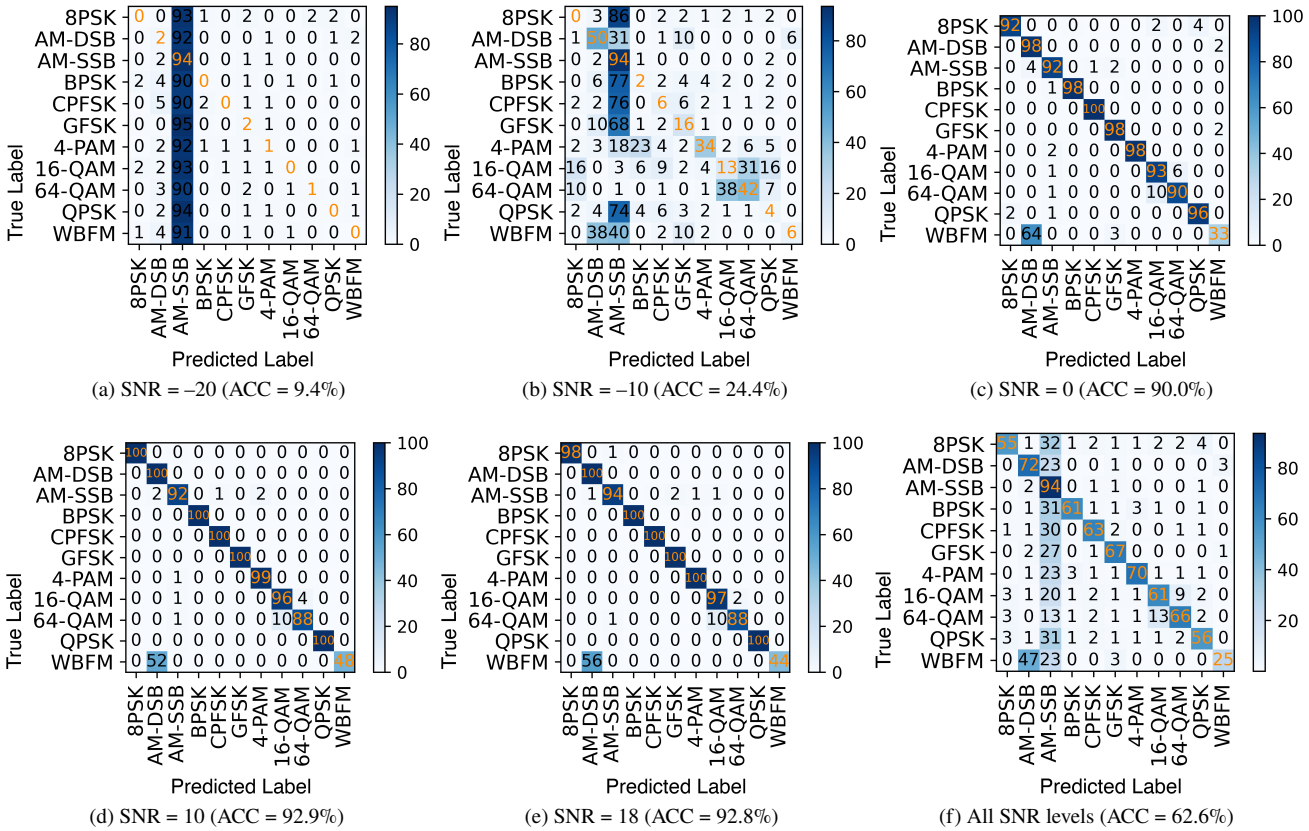
4. Conclusion

We have presented a lightweight four-stream deep learning architecture named TriFusion-Lite for AMR. The proposed framework utilizes a tailored preprocessing pipeline to compress the input signal and extract key features. Our experimental results demonstrate that appropriate compression of *I/Q* signals can significantly improve computational efficiency while maintaining or even enhancing recognition performance. The model then leverages parallel branches to process the signal in the temporal, frequency, and phase-aware domains, thereby effectively capturing a comprehensive signal representation. The model was validated on three benchmark datasets. The results show that it achieves an outstanding balance among classification accuracy, model complexity, and computational efficiency. The low parameter count and high computational throughput of the model make it a highly suitable choice for real-time deployment on resource-constrained edge devices, such as those in 5G/6G



(a) Accuracy vs. epoch (b) Loss vs. epoch

Fig. 7. Training and validation curves of TriFusion-Lite on RML2016.10a.



(a) SNR = –20 (ACC = 9.4%) (b) SNR = –10 (ACC = 24.4%) (c) SNR = 0 (ACC = 90.0%)

(d) SNR = 10 (ACC = 92.9%) (e) SNR = 18 (ACC = 92.8%) (f) All SNR levels (ACC = 62.6%)

Fig. 8. Confusion matrices of TriFusion-Lite across different SNR levels (Performance on RML2016.10a).

networks and IoT ecosystems. Future research will evaluate TriFusion-Lite on non-RadioML datasets (e.g., HisarMod2019 or SDR-captured signals) to verify its robustness in non-simulated environments. Future work will explore the integration of TriFusion-Lite into a multimodal foundation model framework, enabling knowledge transfer across different RF tasks (e.g., signal detection, and emitter identification) through shared representations and multi-task learning.

Acknowledgments

This work was supported in part by the Chengdu Technological University Research Fund under Grant Nos. 2550104 (Rujiao CHENG) and 2024ZR016 (Juan SU).

References

[1] DOBRE, O. A., ABDI, A., BAR-NESS, Y., et al. Survey of automatic modulation classification techniques: Classical approaches and new trends. *IET Communications*, 2007, vol. 1, no. 2, p. 137–156. DOI: 10.1049/iet-com:20050176

[2] WEI, W., MENDEL, J. M. Maximum-likelihood classification for digital amplitude-phase modulations. *IEEE Transactions on Communications*, 2000, vol. 48, no. 2, p. 189–193. DOI: 10.1109/26.823550

[3] HUANG, S., YAO, Y., WEI, Z., et al. Automatic modulation classification of overlapped sources using multiple cumulants. *IEEE Transactions on Vehicular Technology*, 2017, vol. 66, no. 7, p. 6089–6101. DOI: 10.1109/TVT.2016.2636324

[4] GARDNER, W. A. Spectral correlation of modulated signals: Part I – analog modulation. *IEEE Transactions on Communications*, 1987, vol. 35, no. 6, p. 584–594. DOI: 10.1109/TCOM.1987.1096820

[5] KHURANA, L., CHAUHAN, A., NAVED, M., et al. Speech recognition with deep learning. *Journal of Physics: Conference Series*, 2021, vol. 1854, no. 1, p. 1–6. DOI: 10.1088/1742-6596/1854/1/012047

[6] HUIXIAN, J. The analysis of plants image recognition based on deep learning and artificial neural network. *IEEE Access*, 2020, vol. 8, p. 68828–68841. DOI: 10.1109/ACCESS.2020.2986946

[7] O'SHEA, T. J., CORGAN, J., CLANCY, T. C. Convolutional radio modulation recognition networks. In *Proceedings of the International Conference on Engineering Applications of Neural Networks (EANN)*. Aberdeen (UK), 2016, p. 213–226. DOI: 10.1007/978-3-319-44188-7_16

[8] RAJENDRAN, S., MEERT, W., GIUSTINIANO, D., et al. Deep learning models for wireless signal classification with distributed low-cost spectrum sensors. *IEEE Transactions on Cognitive Communications and Networking*, 2018, vol. 4, no. 3, p. 433–445. DOI: 10.1109/TCCN.2018.2835460

[9] O'SHEA, T. J., ROY, T., CLANCY, T. C. Over-the-air deep learning based radio signal classification. *IEEE Journal of Selected Topics in Signal Processing*, 2018, vol. 12, no. 1, p. 168–179. DOI: 10.1109/JSTSP.2018.2797022

[10] LIU, X., YANG, D., EL GAMAL, A. Deep neural network architectures for modulation classification. In *Proceedings of the 51st Asilomar Conference on Signals, Systems, and Computers*. Pacific Grove (CA, USA), 2017, p. 915–919. DOI: 10.1109/ACSSC.2017.8335483

[11] CHENG, R., CHEN, Q., HUANG, M. Automatic modulation recognition using deep CVCNN-LSTM architecture. *Alexandria Engineering Journal*, 2024, vol. 104, p. 162–170. DOI: 10.1016/j.aej.2024.06.008

[12] XIAO, C., YANG, S., FENG, Z. Complex-valued depthwise separable convolutional neural network for automatic modulation classification. *IEEE Transactions on Instrumentation and Measurement*, 2023, vol. 72, p. 1–10. DOI: 10.1109/TIM.2023.3298657

[13] XU, J., WU, C., YING, S., et al. The performance analysis of complex-valued neural network in radio signal recognition. *IEEE Access*, 2022, vol. 10, p. 48708–48718. DOI: 10.1109/ACCESS.2022.3171856

[14] FAN, M., PENG, C., WU, L., et al. Automatic modulation classification: A deep learning enabled approach. *IEEE Transactions on Vehicular Technology*, 2018, vol. 67, no. 11, p. 10760–10772. DOI: 10.1109/TVT.2018.2868698

[15] KARRA, K., KUZDEBA, S., PETERSEN, J. Modulation recognition using hierarchical deep neural networks. In *Proceedings of the IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*. Baltimore (MD, USA), 2017, p. 1–3. DOI: 10.1109/DySPAN.2017.7920746

[16] HAN, J., YU, Z., YANG, J. Multimodal attention-based deep learning for automatic modulation classification. *Frontiers in Energy Research*, 2022, vol. 10, p. 1–10. DOI: 10.3389/fenrg.2022.1041862

[17] XU, J., LUO, C., PARR, G., et al. A spatiotemporal multi-channel learning framework for automatic modulation recognition. *IEEE Wireless Communications Letters*, 2020, vol. 9, no. 10, p. 1629–1632. DOI: 10.1109/LWC.2020.2999453

[18] ZHANG, W., XUE, K., YAO, A., et al. Automatic modulation recognition based on multimodal information processing: A new approach and application. *Electronics*, 2024, vol. 13, no. 22, p. 1–18. DOI: 10.3390/electronics13224568

[19] YI, Z., MENG, H., GAO, L., et al. Efficient convolutional dual-attention transformer for automatic modulation recognition. *Applied Intelligence*, 2025, vol. 55, no. 3, p. 231–244. DOI: 10.1007/s10489-024-06202-6

[20] ZHANG, L., HAO, X., LIU, Q., et al. Automatic modulation recognition of radio frequency proximity sensor signals based on adaptive relational graph attention network. *IEEE Transactions on Instrumentation and Measurement*, 2025, vol. 74, p. 1–13. DOI: 10.1109/TIM.2025.3552477

[21] WANG, G., HAO, X., TIAN, C. DiDbKDT: A cross-SNR modulation recognition scheme with dual-input and dual-branch based on knowledge distillation for IQ-AP data. *IEEE Access*, 2025, vol. 13, p. 143414–143428. DOI: 10.1109/ACCESS.2025.3598978

[22] POLAK, L., TURAK, S., SOTNER, R., et al. Exploring deep learning architectures for RF signal classification. In *Proceedings of the 35th International Conference on Radioelektronika (RADIOELEKTRONIKA)*. 2025, p. 1–6. DOI: 10.1109/RADIOELEKTRONIKA65656.2025.11008396

[23] ZHANG, J., WANG, T., FENG, Z., et al. Toward automatic modulation classification with adaptive wavelet network. *IEEE Transactions on Cognitive Communications and Networking*, 2023, vol. 9, no. 3, p. 549–563. DOI: 10.1109/TCCN.2023.3252580

[24] WANG, M., FANG, S., FAN, Y., et al. An ultra-lightweight neural network for automatic modulation classification in drone communications. *Scientific Reports*, 2024, vol. 14, article 21540. DOI: 10.1038/s41598-024-72867-1

[25] CHEN, T., LIU, K., HUANG, Q. EET-MoCo: An efficient embedding transformer with momentum contrast learning for automatic modulation recognition. *IEEE Transactions on Cognitive Communications and Networking*, 2025, vol. 9, p. 1–9. DOI: 10.1109/TCCN.2025.3541732

[26] SARMANBETOV, S., NURGALIYEV, M., ZHOLAMANOV, B., et al. Novel filtering and regeneration technique with statistical feature extraction and machine learning for automatic modulation classification. *Digital Signal Processing*, 2024, vol. 155, p.1–12. DOI: 10.1016/j.dsp.2024.104744

[27] LIU, X., SONG, Y., ZHU, J., et al. An efficient deep learning model for automatic modulation classification. *Radioengineering*, 2024, vol. 33, no. 4, p. 713–720. DOI: 10.13164/re.2024.0713

[28] KRZYSTON, J., BHATTACHARJEA, R., STARK, A. Complex-valued convolutions for modulation recognition using deep learning. In *Proceedings of the IEEE International Conference on Communications Workshops (ICC Workshops)*. Dublin (Ireland), 2020, p. 1–6. DOI: 10.1109/ICCWorkshops49005.2020.9145469

[29] ZHANG, F., LUO, C., XU, J., et al. An efficient deep learning model for automatic modulation recognition based on parameter estimation and transformation. *IEEE Communications Letters*, 2021, vol. 25, no. 10, p. 3287–3290. DOI: 10.1109/LCOMM.2021.3102656

[30] QU, Y., LU, Z., ZENG, R., et al. Enhancing automatic modulation recognition through robust global feature extraction. *IEEE Transactions on Vehicular Technology*, 2024, vol. 73, p. 1–11. DOI: 10.1109/TVT.2024.3486079

About the Authors . . .

Rujiao CHENG (corresponding author) was born in 1991. She received the M.Sc. degree in Electronic and Communication Engineering from East China Normal University in 2017. Her research interests include signal processing and pattern recognition.

Juan SU was born in 1989. She received the M.Sc. degree in Circuits and Systems from Sichuan University in 2016. Her research interests include signal processing, modulation recognition, and deep learning.

Min HUANG was born in 1992. She received the M.Sc. degree from Sichuan Normal University in 2024. Her research interests include time-frequency analysis, lightweight neural networks, and feature fusion.