AMR-LLM: Knowledge-Enhanced Multi-Modal Automatic Modulation Recognition via Large Language Models

Shen Hu[1,2], **Yuhua Qian**[1,2*], **Xinyan Liang**[1,2], **Zikun Jin**[1], **Jiaqian Zhang**[1] and **Jiangfeng Zhang**[1]
[1]Institute of Big Data Science and Industry, Key Laboratory of Evolutionary Science Intelligence of Shanxi Province, Shanxi University, Taiyuan, China
[2]School of Artificial Intelligence, Shanxi University, Datong, China
shenhu365@outlook.com, jinchengqyh@126.com, liangxinyan48@163.com, jinzikun@163.com, zhangjiaqian1@sxu.edu.cn, zjf_8099@163.com

Abstract

Existing multi-modal automatic modulation recognition (AMR) methods primarily focus on exploiting multi-view representations of raw signal data to improve performance, but still struggle to effectively model and exploit high-level human prior knowledge. Although recent studies attempt to introduce large language models (LLMs) to integrate the textual modality, most of them merely treat LLMs as feature extractors, without further exploiting the LLM's potential. To this end, we propose a knowledge-enhanced multi-modal automatic modulation recognition framework based on large language model (AMR-LLM). Specifically, we first specially design an AMR instruction construction mechanism based on knowledge to activate the LLM's potential for signal perception, which designates the LLM as a signal domain expert, introduces human prior knowledge as a supplement, and exploits the unique physical information of each data sample as connection guidance. Then, to enable the LLM to directly perceive digital signals, we introduce a progressive multi-modal fusion strategy mapping signal features into the space of LLMs. Experimental results on multiple benchmark datasets demonstrate that AMR-LLM achieves approximately 5% higher accuracy than current state-of-the-art methods. Moreover, it achieves the current domain performance with only 30% of the data.

1 Introduction

With the development of cognitive radio, the electromagnetic spectrum environment has become increasingly congested and complex. As the "eyes" of cognitive radio, automatic modulation recognition (AMR) is critical for spectrum monitoring, interference identification, and signal demodulation. In real-world implementation, it is still a challenging topic to achieve strong AMR performance under harsh conditions, such as low signal-to-noise ratio (SNR), non-line-of-sight (NLOS) propagation, and dynamic channel fading.

*Corresponding author: jinchengqyh@126.com

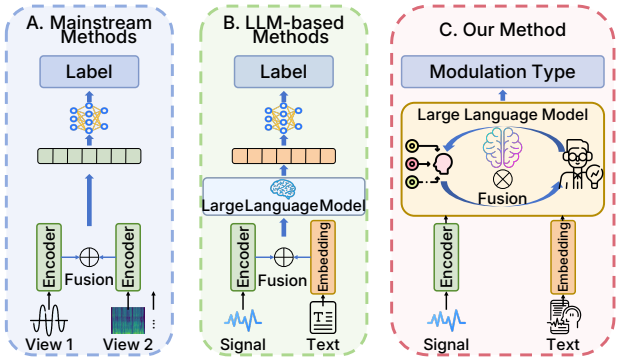

Figure 1: (a) Mainstream methods (multi-view). (b) LLM-based methods treat LLMs as feature extractors. (c) Our knowledge-enhanced framework activates the LLM as an AMR expert and enables it to perceive signals. Data sourced from (a) [Shao *et al.*, 2025] (b) [Chen *et al.*, 2025].

Current methods primarily rely on deep learning models to learn discriminative features from raw signal data or time-frequency representations. Although such mainstream methods achieve impressive performance on certain datasets, they are unable to flexibly model and utilize one essential dimension of information, i.e., high-level human prior knowledge, especially when the available evidence is noisy, incomplete, or heterogeneous [Wen *et al.*, 2026; Guo *et al.*, 2025]. As an intuitive example, in image segmentation or recognition, textual prior knowledge of the target can guide the model's attention toward semantically relevant regions and thereby help localize the target [Zhang *et al.*, 2023c], even under occlusion or ambiguity. This idea of knowledge-guided perception is analogous to the way human experts reason in signal analysis. Experts typically make comprehensive judgments by integrating complementary evidence and domain knowledge, such as modulation theories, constellation distribution laws, and channel characteristics [Liang *et al.*, 2025a]. This is also consistent with recent studies showing that expert knowledge constraints and structure-aware multi-perspective feature modeling can improve representation robustness under heterogeneous evidence [Wu *et al.*, 2025; Liang *et al.*, 2025c; Yang *et al.*, 2025].

Nevertheless, even with recent progress in dynamic or graph-guided multi-view learning [Ren *et al.*, 2024; Chao *et*

al., 2025], abstract knowledge still imposes a major challenge to conventional deep neural networks to encode and exploit it directly. The chronic bottleneck still persists in how to allow models to understand and exploit such high-level priors similar to humans. With the advent of Large Language Models (LLMs), a fresh opportunity emerges to solve the bottleneck. As knowledge stores covering extensive human knowledge, LLMs enjoy a strong inherent power of abstract understanding and inference [Wei *et al.*, 2022]. Recent large language model-based work for automatic modulation recognition (RadioLLM [Chen *et al.*, 2025]) has made valuable initial explorations, primarily treating the large language model as a feature extractor. We observe that there remains an opportunity to more fully leverage the LLM's original capabilities, including deep generative inference and its rich general knowledge base [Luo *et al.*, 2026]. We argue that a knowledge-enhanced automatic modulation recognition paradigm should enable the large language model to act as a domain expert, inferring from signals with the help of knowledge.

Further focus on LLMs leads us to an interesting thought: The large language model remains unchanged across the globe, but the output for different users changes greatly. The performance disparity between a child and an AMR domain expert using the exact same large language model is huge, and the only difference is the amount of prior knowledge encoded in the prompt.

To this end, we propose AMR-LLM, a knowledge-enhanced multi-modal AMR framework based on LLMs. AMR-LLM enriches conventional feature-based recognition with knowledge-guided inference, enabling signal features, expert priors, and reasoning capabilities to complement each other.

Specifically, we first specially design an AMR instruction construction mechanism based on knowledge to activate the LLM's potential for signal perception. In this mechanism, the prompt plays two roles: (a) Complementarity: It provides general AMR expert knowledge (e.g., modulation is in amplitude, phase, and frequency; data distribution changes across different SNRs) to complement signal features and evoke the implicit AMR expert knowledge encoded in the large language model. (b) Guiding Role: This is the link between signal and language; we take the special physical attributes of each sample and infuse them into the prompt, which serves as a guide for the signal connection.

At this step, the large language model faces the problem of perceiving digital signals. Considering the intrinsic heterogeneity between signal data and textual knowledge, we propose a Progressive Multi-modal Fusion Strategy. With the dynamic injection of signal features into multiple levels of the large language model and the assembly of the anchor information in the prompt, guide the mapping of signal features into the space of LLMs, enabling the large language model to perceive the signal.

The main contributions of this work are as follows:

- We propose AMR-LLM, a knowledge-enhanced recognition framework based on large language model, which consists of three key modules: physical perception modelling, knowledge enhancement design, and synergy of knowledge and perception. This framework provides a new reference for future multi-modal AMR.
- We propose an instruction construction method with a bi-facet function. This method can not only enhance and activate the LLM's AMR knowledge and potential, but also mitigate the difficulty of signal perception for the LLM through the combined "anchor" effect of physical features and a progressive fusion strategy.
- Experimental results on multiple benchmark datasets demonstrate that AMR-LLM achieves approximately 5% higher accuracy than current state-of-the-art (SOTA) methods. Moreover, it achieves current domain performance with only 30% of the data.

2 Related Work

2.1 Robust Multi-View Methods

Recent research has therefore moved from single-view models toward multi-view representation learning, as robust recognition under complex data conditions often relies on reliable evaluation and stable feature representations [Qiu *et al.*, 2024; Qiu *et al.*, 2025a]. For example, IQFormer [Shao *et al.*, 2025] adopts a Transformer-based architecture to integrate spatial-temporal I/Q features with time-frequency distributions. At the representation level, transform-domain features and dynamic alignment mechanisms have also been used to exploit complementary signal information in complex signal scenarios [Jin *et al.*, 2025; Wang *et al.*, 2025]. These studies suggest that robustness depends not only on effective heterogeneous feature fusion, but also on modeling latent relationships and interactions among features [Qiu *et al.*, 2025b; Jin *et al.*, 2026]. Complementing these architectural advances, recent works such as FR-AMC [Liang *et al.*, 2025b] further improve model robustness from the perspective of decision-boundary optimization by using fuzzy regularization.

2.2 LLM-based Methods

The rise of large language models (LLMs) has led to a shift from purely discriminative to generative frameworks in wireless communications. To tackle the problem of recognizing unseen modulations, Plug-and-Play AMC [Rostami *et al.*, 2025b] and DiSC-AMC [Rostami *et al.*, 2025a] propose an In-Context Learning (ICL) framework: they convert quantitative signal statistics into textual prompts, and then use frozen LLMs to perform "training-free" classification by logical deduction. At the same time, ZSAMR-VLM [Zhao *et al.*, 2025] explores cross-modal alignment by using vision-language models to map signal spectrograms into linguistic descriptions for zero-shot recognition. However, such methods have not achieved effective recognition by combining signal and text modalities under extremely low SNR. Recent methods like RadioLLM [Chen *et al.*, 2025] have pioneered the alignment of radio signals and LLM embeddings via Hybrid Prompt and Token Reprogramming (HPTR). However, RadioLLM mainly uses the LLM as a reprogrammed feature extractor. In this work, we aim to activate the LLM as a domain expert that can utilize prior knowledge and signal data to infer the modulation type.

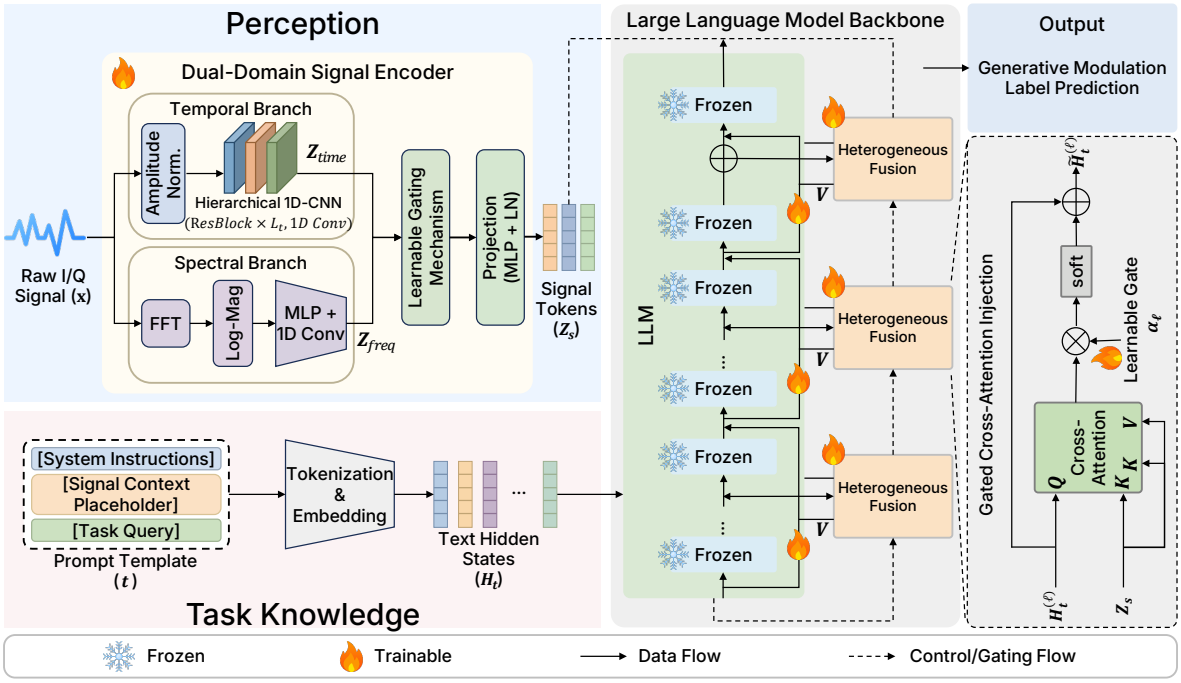

Figure 2: Overall architecture of AMR-LLM. The framework consists of three key modules: (1) Physical Perception Modeling for LLM; (2) Knowledge Enhancement Design; (3) Synergy of Knowledge and Perception.

3 Methodology

In this section, we present the proposed knowledge-enhanced multi-modal automatic modulation recognition framework based on large language model (AMR-LLM). To effectively inject domain knowledge, we adopt LLMs as the carrier. However, LLMs inherently lack the capability to perceive physical signals. Therefore, we design three key modules: (1) physical perception modelling, (2) Knowledge enhancement design, and (3) synergy of knowledge and perception. The overall architecture is illustrated in Fig. 2.

3.1 Problem Formulation

Denote the set of labeled raw signals $\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_{i=1}^{N}$, where each sample $\mathbf{x} \in \mathbb{R}^{2\times L}$ is a complex-valued In-phase and Quadrature (IQ) sequence of length L and $\mathbf{x} = [\mathbf{x}_I; \mathbf{x}_Q]$. The goal is to predict the modulation category $y \in \mathcal{Y}$ of signal $\mathbf{x}$, where $|\mathcal{Y}| = C$.

Unlike the prevalent discriminative models that estimate $P(y|\mathbf{x})$ directly, we recast AMR as a knowledge-enhanced conditional generation task. Specifically, we construct a structured textual instruction $\mathbf{t}$ that represents the general domain knowledge and the instance-specific physical prompting. The objective is to learn the parameters θ such that the probability of the target label sequence tokens is maximized, given the physical observation and the expert instruction:

$$y^* = \arg\max_{y\in\mathcal{Y}} P_\theta(y|\mathbf{x}, \mathbf{t}). \tag{1}$$

This formulation allows the model to leverage the vast internalized knowledge of LLMs.

3.2 Physical Perception Modeling

The native format of input accepted by LLMs is textual language. However, the raw data containing relatively authentic physical information is in the form of digital signals. As a first step, we aim to enable large language models to perceive physical signals.

Each sample is not merely a set of numerical values but a physical law projected in the time-frequency domain. To map raw IQ samples into the space of LLMs, relying on a single domain is insufficient to well reflect the complex physical characteristics of RF signals. We propose a dual-domain signal encoder($\mathcal{E}_s$), which perceives the dynamics of signals via two complementary branches: a temporal branch and a spectral branch. We assume the raw IQ signal is $\mathbf{x}$. We first apply instance normalization to flatten the power scale and thus ensure numerical stability, and then process through two parallel branches.

Temporal Stream: To model the instantaneous phase shifts and amplitude variations, we use a hierarchical residual CNN architecture with minimal downsampling. The temporal representation is provided as:

$$\mathbf{Z}_{time} = \text{Pool}_{N_s}(\text{Conv1D}(\mathbf{h}^{(L_t)})) \in \mathbb{R}^{N_s\times d_s}, \tag{2}$$

where d_s is the hidden dimension of the signal encoder, the branch only performs $2\times$ overall downsampling, maintaining the temporal resolution required to differentiate modulations.

Spectral Stream: We use the spectral branch to explicitly capture frequency-domain properties, such as bandwidth, carrier offset, and spectral shape, which are key for discriminating frequency-shift keying (FSK) modulations and noise levels. We perform the Fast Fourier Transform (FFT) and then

compute the log magnitude. The log transform flattens the dynamic range of the spectrum, allowing the network to learn both the strong carrier components and the fine spectral morphology. The spectral representation is then retrieved from:

$$\mathbf{Z}_{freq} = \text{Pool}_{N_s}(\text{Conv1D}(\text{MLP}(\mathbf{S}))) \in \mathbb{R}^{N_s \times d_s}. \quad (3)$$

Instead of simply concatenating the two branches, we dynamically fuse them with a learnable gating mechanism to appropriately weigh the contributions from both branches according to signal characteristics:

$$\mathbf{Z}_{fused} = \sigma(\alpha_g) \cdot \mathbf{Z}_{time} + (1 - \sigma(\alpha_g)) \cdot \mathbf{Z}_{freq}, \quad (4)$$

where α_g is a learnable scalar parameter and $\sigma(\cdot)$ denotes the sigmoid function. During training, our model learns to focus more dynamically on the more informative domain for each modulation type. The fused features are further projected into the LLM hidden dimension d via a two-layer MLP with LayerNorm (LN):

$$\mathbf{Z}_s = \text{LN}(\text{MLP}(\mathbf{Z}_{fused})) \in \mathbb{R}^{N_s \times d}, \quad (5)$$

where N_s denotes the number of signal tokens obtained by adaptive average pooling. This projection ensures that the signal representations lie on a manifold compatible with the LLM's embedding space.

3.3 Knowledge Enhancement Design

The capability of an LLM to function as a domain expert depends critically on the quality of its context. We argue that the performance disparity between a generalist model and a signal expert lies in the activation of prior knowledge. To this end, we design a specialized instruction construction mechanism $\mathcal{T}(\cdot)$ that generates a compound prompt $\mathbf{t}$ with a bi-facet function:

$$\mathbf{t} = \mathcal{T}(\mathbf{x}_{meta}) = [\mathbf{t}_{exp}; \mathbf{t}_{gui}], \quad (6)$$

where $\mathbf{x}_{meta}$ represents the meta information of the signal.

The expert Role ($\mathbf{t}_{exp}$): This component injects general AMR expert knowledge that is implicit in signal processing theory but absent in raw signal data, e.g., modulation principles and environmental context (detailed prompt template shown in Fig. 3). More importantly, this component serves to enhance and activate the implicit automatic modulation recognition knowledge and capabilities within the LLM, enabling it to perform expert-level inference.

The guidance Role ($\mathbf{t}_{gui}$): One of the main challenges in multi-modal automatic modulation recognition is that raw signals, after embedding projection, may appear as "semantic noise" to the LLM. To mitigate this phenomenon, we propose adding instance-specific physical anchors to the prompt, such as high-dimensional physical manifold descriptions (topological geometric features, time-domain dynamics analysis, cyclostationary and spectral fingerprints) and statistical features (standard deviation of instantaneous amplitude, peak-to-average power ratio, kurtosis, fourth-order cumulants), etc. (detailed prompt template shown in Fig. 3). However, the content that can be added to the prompt is always limited, so it needs to be combined with the powerful intrinsic potential of the LLM. These "cognitive bridges" aim to guide the mapping of the abstract signal embeddings $\mathbf{Z}_s$ into the LLM's space.

The prompt input must strictly prevent data leakage, with no ground truth labels or SNR information allowed. The prompt construction template is illustrated in Fig. 3. The prompt length is limited to a maximum of 4096 tokens, with a reserved `<|signal|>` placeholder for signal embeddings. The instance-specific anchor information in the prompt is precomputed and extracted using traditional digital signal processing algorithms before training and incorporated into the prompt in text form. The preprocessing time is minimal, requiring only about 1 minute for 100,000 samples.

Prompt Template for AMR

System Role Instruction
Role Spec: You are a domain expert in Automatic Modulation Recognition (AMR) and Signal Processing.
Goal: Your task is to analyze raw signal representations, high-dimensional physical manifold features, and statistical fingerprints to identify the modulation type.
Background Knowledge: The signal modulations include a wide range of types: Amplitude, Phase, Frequency, Quadrature, etc. Stochastic SNR fluctuations and complex channel impairments are considered. Physical manifold descriptions provide geometric insights into the signal's behavior in high-dimensional space...

User Input Prompt
Physical Representation Input
`<|signal|>`
High-Dimensional Physical Manifold Descriptions
Topological Geometric Features:
Manifold Dimension: *[computed value]*
Time-Domain Dynamics Analysis:
Trajectory Stability: *[computed value]*
Spectral & Cyclostationary Fingerprints:
Spectral Flatness: *[computed value]*...
Statistical Fingerprint
High-Order Statistics: Sigma_AA, C42_Abs, Kurtosis (computed value)
Power & Envelope Statistics: PAPR (dB), Max_Amp, Min_Amp, Mean_Amp (computed value)
Specific Instruction: Based on the physical manifold descriptions and statistical metrics above, reason and predict the modulation type.

Figure 3: The prompt instruction template integrating expert knowledge and instance-specific physical information.

The final recognition result is generated by the LLM based on two complementary inputs: the signal tokens $\mathbf{Z}_s$ representing physical observations, and the textual prompt $\mathbf{t}$ providing domain knowledge and guidance. The textual input is first tokenized into embeddings $\mathbf{H}_t \in \mathbb{R}^{N_t \times d}$, and then the LLM generates the prediction autoregressively:

$$y^* = \arg\max_y \prod_i P(w_i | w_{<i}, \mathbf{Z}_s, \mathbf{H}_t). \quad (7)$$

During inference, to improve prediction speed, the LLM is

constrained to output only modulation type prediction-related information, with the maximum output length set to 50 tokens. The predicted labels are then extracted from the text-form prediction results for each sample to calculate various performance metrics.

3.4 Synergy of Knowledge and Perception

Directly concatenating signal tokens $\mathbf{Z}_s$ with text embeddings $\mathbf{H}_t$ will cause LLM to not pay attention to the signal tokens because of the mismatch between continuous signal representation and discrete text embeddings. In order to propose deep synergy, we propose a Progressive Multi-Modal Fusion Strategy. Instead of injecting the signal only once, we progressively impose the physical signal into multiple layers of the LLM for hierarchical refinement.

More concretely, we adopt a Gated Cross-Attention where the internal text states of the LLM are treated as queries asking to confirm from the signal evidence (Keys/Values). For a given transformer layer ℓ, let $\mathbf{H}_t^{(\ell)}$ be the hidden states for the text instructions. We are mapping the signal feature to the current level using the layer-wise projection $\mathbf{W}_K^{(\ell)}$ and $\mathbf{W}_V^{(\ell)}$:

$$\mathbf{Q} = \mathbf{H}_t^{(\ell)}\mathbf{W}_Q^{(\ell)}, \quad \mathbf{K} = \mathbf{Z}_s\mathbf{W}_K^{(\ell)}, \quad \mathbf{V} = \mathbf{Z}_s\mathbf{W}_V^{(\ell)}. \tag{8}$$

Next, the multi-head attention (MHA) is applied to enable the model to jointly attend to information from different representation subspaces:

$$\mathbf{O}_{cross}^{(\ell)} = \text{MHA}(\mathbf{Q}, \mathbf{K}, \mathbf{V})\mathbf{W}_O^{(\ell)}, \tag{9}$$

where $\mathbf{W}_O^{(\ell)} \in \mathbb{R}^{d\times d}$ is the output projection matrix that maps the output back to the LLM's representation space. To ensure stable modality fusion, we introduce a learnable gate α_ℓ to control how much signal information is injected at each layer. The updated hidden state is computed as:

$$\tilde{\mathbf{H}}_t^{(\ell)} = \mathbf{H}_t^{(\ell)} + g(\alpha_\ell) \cdot \mathbf{O}_{cross}^{(\ell)}, \tag{10}$$

where $g(\alpha_\ell)$ is a soft gating function that constrains the injection weight to $[0, 1]$. The learnable parameter α_ℓ modulates the injection strength, i.e., the proportion with which the model incorporates the signal information into the LLM at each transformer layer. We inject multiple cross-attention modules into multiple transformer layers. This way, the signal information is modulated across different LLM layers. During training, the gate parameters α_ℓ and the projection matrices $\mathbf{W}_K^{(\ell)}, \mathbf{W}_V^{(\ell)}$, and $\mathbf{W}_O^{(\ell)}$ are jointly optimized, enabling the model to learn the fusion process adaptively in a data-driven manner.

3.5 Two-Stage Training Strategy

To explicitly achieve robust physical representations and multimodal alignment, we propose a two-stage optimization strategy. In this work, we employ Llama-3-8B-Instruct as the backbone Large Language Model (LLM).

Stage 1: Physical Encoder Pre-training. We freeze the LLM and train the perception module only with a lightweight classification head via cross-entropy loss:

$$\mathcal{L}_{Stage1} = -\sum_{c=1}^{C} y_c \log(\hat{y}_c), \tag{11}$$

to force the encoder to learn discriminative signals with explicit alignment to modulation categories, thus learning a strong physical foundation.

Stage 2: Knowledge-Enhanced Instruction Tuning. We remove the classification head and connect the encoder into the LLM using the progressive fusion modules. We apply Low-Rank Adaptation [Hu *et al.*, 2022] to the LLM parameters. We optimize the whole system using the next-token prediction objective:

$$\mathcal{L}_{Stage2} = -\mathbb{E}_{(\mathbf{Z}_s,\mathbf{t},\mathbf{w})\sim\mathcal{D}}\left[\sum_{i=1}^{T} \log P_\theta(w_i|w_{<i}, \mathbf{Z}_s, \mathbf{t})\right] \tag{12}$$

4 Experiments

4.1 Datasets

To perform an exhaustive assessment across varying signal complexities and channel conditions, we utilize multiple publicly available and widely-used benchmark datasets in this field: RML2016.10a, RML2016.10b [O'shea and West, 2016], RML2016.04c [O'Shea *et al.*, 2016], and RML2022 [Sathyanarayanan *et al.*, 2023]. These benchmarks cover a diverse spectrum of modulation schemes and signal-to-noise ratios (SNRs). The dataset was split into training, validation, and test sets with a ratio of 6:2:2. To ensure statistical consistency, we applied a strict stratified random sampling strategy for dataset partitioning. Specifically, sampling was performed uniformly across every distinct SNR level and modulation category, thereby preserving the intrinsic distribution of signal characteristics throughout the experimental subsets.

4.2 Training Setup and Evaluation Metrics

All experiments were conducted on NVIDIA A100 GPUs using PyTorch 2.6.0. Stage I training used a learning rate of 5.0×10^{-4} with 30 epochs. Stage II used a learning rate of 1.0×10^{-4} with 30 epochs, cosine learning rate scheduler with warmup ratio of 0.1. For LoRA configuration, we set rank to 32, alpha to 64, and dropout to 0.1. The signal injection was applied at layers $\{0, 4, 8, 12, 16, 20, 24, 28\}$ with 8 attention heads. We employed four complementary metrics to evaluate classification performance: Overall Accuracy (ACC), Cohen's Kappa (κ), Macro F1-Score, and Matthews Correlation Coefficient (MCC), these metrics collectively assess both overall correctness.

4.3 Performance Analysis

We compare AMR-LLM against seven representative methods in Table 1, including five mainstream methods (AWN [Zhang *et al.*, 2023b], AMC-Net [Zhang *et al.*, 2023a], TLDNN [Qu *et al.*, 2024], SigDA [Wang *et al.*, 2024], IQFormer [Shao *et al.*, 2025]) and two LLM-based methods (FSCA [Hu *et al.*, 2025], RadioLLM [Chen *et al.*, 2025]), where RadioLLM is specifically designed for AMR tasks and FSCA is a representative method from the time series domain. The experimental results of comparison methods were directly reproduced using the same 6:2:2 data split ratio as ours, while keeping the training hyperparameters consistent with their original implementations.

Dataset	Methods	Pub./Year	Overall Accuracy	Kappa	Macro F1-Score	MCC
RML2016.10a	AWN	IEEE TCCN/2023	61.96	0.5816	0.6408	0.5977
	AMC-Net	IEEE ICASSP/2023	62.45	0.5869	0.6462	0.6043
	TLDNN	IEEE TVT/2024	61.04	0.5715	0.6316	0.5875
	SigDA	IEEE TWC/2024	62.82	0.5911	0.6495	0.6091
	IQFormer	IEEE TCCN/2025	<u>63.92</u>	<u>0.6032</u>	<u>0.6629</u>	<u>0.6205</u>
	FSCA*	ICLR/2025	55.53	0.5108	0.5639	0.5200
	RadioLLM	ARXIV/2025	58.10	0.5391	-	-
	AMR-LLM (Ours)		**70.04**	**0.6704**	**0.7013**	**0.6705**
	Improvement		6.12↑	0.0673↑	0.0384↑	0.0500↑
RML2016.10b	AWN	IEEE TCCN/2023	64.27	0.6030	0.6447	0.6072
	AMC-Net	IEEE ICASSP/2023	65.07	0.6119	0.6517	0.6163
	TLDNN	IEEE TVT/2024	65.35	0.6151	0.6575	0.6194
	SigDA	IEEE TWC/2024	65.00	0.6111	<u>0.6654</u>	<u>0.6220</u>
	IQFormer	IEEE TCCN/2025	<u>65.37</u>	<u>0.6152</u>	0.6528	0.6180
	FSCA*	ICLR/2025	59.95	0.5550	0.6028	0.5620
	RadioLLM	ARXIV/2025	58.35	0.5372	-	-
	AMR-LLM (Ours)		**70.71**	**0.6746**	**0.7101**	**0.6750**
	Improvement		5.34↑	0.0594↑	0.0526↑	0.0556↑
RML2016.04c	AWN	IEEE TCCN/2023	71.98	0.6821	0.7375	0.6834
	AMC-Net	IEEE ICASSP/2023	66.36	0.6183	0.6560	0.6212
	TLDNN	IEEE TVT/2024	65.23	0.6098	0.6663	0.6204
	SigDA	IEEE TWC/2024	<u>73.20</u>	<u>0.6971</u>	<u>0.7375</u>	<u>0.6977</u>
	IQFormer	IEEE TCCN/2025	72.03	0.6836	0.7309	0.6841
	FSCA*	ICLR/2025	66.95	0.6273	0.6793	0.6274
	RadioLLM	ARXIV/2025	68.19	0.6441	-	-
	AMR-LLM (Ours)		**78.86**	**0.7623**	**0.7653**	**0.7633**
	Improvement		5.66↑	0.0652↑	0.0278↑	0.0656↑
RML2022	AWN	IEEE TCCN/2023	68.06	0.6451	0.6775	0.6457
	AMC-Net	IEEE ICASSP/2023	68.77	0.6530	0.6847	0.6540
	TLDNN	IEEE TVT/2024	67.44	0.6382	0.6713	0.6398
	SigDA	IEEE TWC/2024	69.81	0.6645	<u>0.6981</u>	<u>0.6680</u>
	IQFormer	IEEE TCCN/2025	<u>69.83</u>	<u>0.6648</u>	0.6962	0.6653
	FSCA*	ICLR/2025	61.08	0.5675	0.6055	0.5684
	RadioLLM	ARXIV/2025	59.39	0.5488	-	-
	AMR-LLM (Ours)		**74.86**	**0.7195**	**0.7376**	**0.7174**
	Improvement		5.03↑	0.0547↑	0.0395↑	0.0494↑

Table 1: Performance comparison on four benchmark datasets. We compare AMR-LLM against seven representative methods, including five mainstream AMR approaches, one representative time-series foundation model method (FSCA)*, and one LLM-based AMR method (RadioLLM).

AMR-LLM achieved improvements across all metrics on all benchmark datasets. For instance, on the RML2016.10a dataset, the accuracy improved by 6.12%, Cohen's Kappa increased by 0.0673, Macro F1-Score improved by 0.0384, and MCC increased by 0.0500. Consistent improvements were also observed on other datasets, indicating that combining knowledge enhancement with physical perception exhibits substantial performance advantages over existing SOTA methods. Meanwhile, we observed that although prior LLM-based methods already utilized both physical signals and textual information, their performance was still slightly inferior to mainstream methods that relied solely on physical signal data. This may provide a useful reference that both knowledge enhancement and physical perception need to be carefully considered and effectively combined to better unleash the potential of LLMs.

4.4 The Importance of Knowledge and Perception

As shown in Table 1, AMR-LLM has achieved relatively high performance. However, to validate whether knowledge and perception truly contribute to the performance, we conduct

Perception	Knowledge	RML2016a			
		ACC	Kappa	F1	MCC
✗	✓	34.74	0.2821	0.3459	0.2841
✓	✗	57.98	0.5378	0.5938	0.5534
✓	✓	**70.04**	**0.6704**	**0.7013**	**0.6705**

Table 2: Ablation study on the importance of knowledge enhancement and physical perception on RML2016.10a.

experiments using only knowledge enhancement and only physical perception separately on the RML2016.10a dataset. Results are shown in Table 2.

First, we find that using only knowledge enhancement without physical perception yields poor recognition performance, which aligns with human intuition. To further investigate where the performance degrades, we analyze the recognition accuracy across different SNR levels and discover that the model only achieves reasonable recognition capability in high-quality environments (SNR $>$ 10 dB). Second, using only physical perception data without knowledge enhancement, the recognition accuracy is similar to previous LLM-based methods. Finally, when using both components together, the recognition accuracy improves by approximately 5% compared to existing SOTA. This further demonstrates that effectively combining physical perception with knowledge enhancement can better unleash the potential of LLMs.

4.5 Advantage in Low-SNR

AMR-LLM indeed achieves significantly higher overall performance than existing SOTA methods in this field, but we want to understand where exactly the improvements come from. Since existing methods have already achieved satisfactory performance under high-SNR conditions, we examine the dynamics of these methods under low-SNR scenarios.

We compare the recognition accuracy of representative methods across different SNR levels on the RML2016.10a dataset, as shown in Figure 4. The results reveal that AMR-LLM's main improvements occur in challenging low-SNR conditions, with substantial gains in the -20 to -10 dB range. In particular, at the most challenging SNR level of -20 dB, its accuracy improves by approximately 20% compared to existing methods. We believe this may be attributed to the powerful inference capability and fine-grained attention ability of LLMs, which enable the model to capture discriminative information even under low-SNR conditions. This also provides a potential reference direction for weak signal recognition in complex scenarios.

4.6 Data Efficiency Analysis

While achieving high accuracy, we are curious whether introducing LLMs can reduce the dependency on training data and whether comparable performance can be achieved with less data, since large language models inherently contain implicit knowledge and inference capabilities related to AMR.

To address this question, we train AMR-LLM with varying data proportions ranging from 10% to 60% across multiple

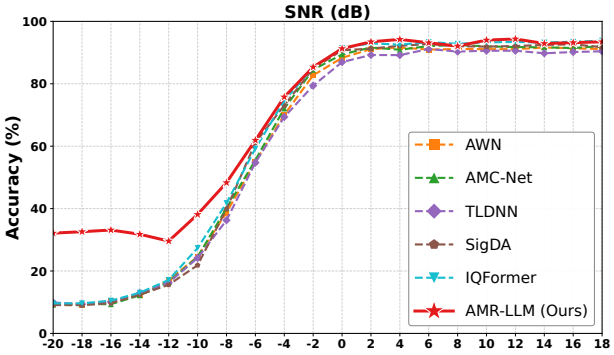

Figure 4: Recognition accuracy comparison across different SNR levels on RML2016.10a.

benchmarks. Figure 5 presents the results. The results reveal that AMR-LLM trained with approximately 30% of the data already achieves performance comparable to current domain SOTA methods. Although AMR-LLM achieves significant performance improvements, it does incur higher computational overhead due to the LLM backbone(0.1 s per sample). However, this approach with less data dependency can serve as a powerful "teacher" model for model distillation or guiding lightweight models, enabling rapid domain adaptation and deployment in resource-constrained scenarios (e.g., data scarcity or prohibitive labeling costs).

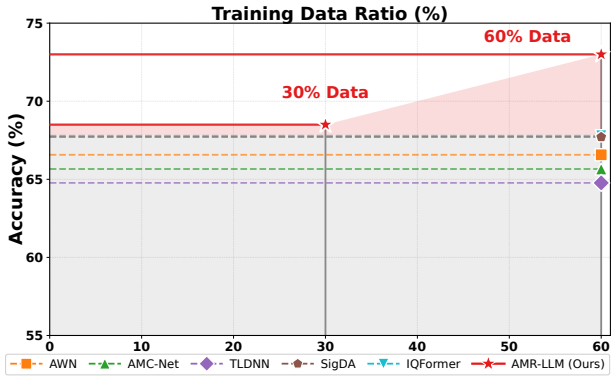

Figure 5: Data efficiency analysis on multiple benchmark datasets.

5 Conclusion

The success of AMR-LLM demonstrates that, in complex environments with challenging recognition tasks but less stringent real-time constraints, effective recognition can be explored through collaboration between large and lightweight models, where lightweight models handle routine samples and large models are invoked for complex cases. Furthermore, this framework can evolve into intelligent agents with perception and reasoning capabilities. By synergizing human prior knowledge, such as expert experience and physical laws, with bottom-up machine perception from signal features, AMR-LLM contributes to the exploration of AI agent systems for signal processing that integrate data sensitivity with logical reasoning capabilities.

Acknowledgements

This work was supported by the Major Program of National Natural Science Foundation of China (T2495251), the Key Program of the National Natural Science Foundation of China (62136005) and the Special Fund for Science and Technology Innovation Teams of Shanxi Province(No.202304051001001).

References

[Chao *et al.*, 2025] Guoqing Chao, Zhenghao Zhang, Lei Meng, Jie Wen, and Dianhui Chu. Federated incomplete multi-view clustering with globally fused graph guidance. In *International Conference on Machine Learning*, pages 7417–7429. PMLR, 2025.

[Chen *et al.*, 2025] Shuai Chen, Yong Zu, Zhixi Feng, Shuyuan Yang, and Mengchang Li. RadioLLM: Introducing large language model into cognitive radio via hybrid prompt and token reprogrammings, 2025.

[Guo *et al.*, 2025] Jipeng Guo, Yanfeng Sun, Xin Ma, Junbin Gao, Yongli Hu, Youqing Wang, and Baocai Yin. Globality meets locality: An anchor graph collaborative learning framework for fast multiview subspace clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 36(6):10213–10227, 2025.

[Hu *et al.*, 2022] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022.

[Hu *et al.*, 2025] Yuxiao Hu, Qian Li, Dongxiao Zhang, Jinyue Yan, and Yuntian Chen. Context-Alignment: Activating and enhancing LLMs capabilities in time series. In *The Thirteenth International Conference on Learning Representations*, 2025.

[Jin *et al.*, 2025] Zikun Jin, Yuhua Qian, Xinyan Liang, and Haijun Geng. A multi-view fusion approach for enhancing speech signals via short-time fractional fourier transform. In *Proceedings of the 34th International Joint Conference on Artificial Intelligence*, pages 5508–5516, 2025.

[Jin *et al.*, 2026] Zikun Jin, Yuhua Qian, Xinyan Liang, Jiaqian Zhang, Jinpeng Yuan, Shen Hu, Haijun Geng, and Honghong Cheng. Signal enhancement via multi-view dynamic representation and alignment-aware fusion. In *AAAI*, pages 22454–22462, 2026.

[Liang *et al.*, 2025a] Xinyan Liang, Shuai Li, Qian Guo, Yuhua Qian, Bingbing Jiang, Tingjin Luo, and Liang Du. Improving evolutionary multi-view classification via eliminating individual fitness bias. In *Advances in Neural Information Processing Systems*, volume 38, pages 143388–143412. Curran Associates, Inc., 2025.

[Liang *et al.*, 2025b] Xinyan Liang, Ruijie Sang, Yuhua Qian, Qian Guo, Feijiang Li, and Liang Du. Robust automatic modulation classification with fuzzy regularization. In *Forty-second International Conference on Machine Learning*, volume 267, pages 37396–37408, 2025.

[Liang *et al.*, 2025c] Xinyan Liang, Shijie Wang, Yuhua Qian, Qian Guo, Liang Du, Bingbing Jiang, Tingjin Luo, and Feijiang Li. Trusted multi-view classification with expert knowledge constraints. In *Proceedings of the 42th International Conference on Machine Learning*, 2025.

[Luo *et al.*, 2026] Haoxiang Luo, Yu Yan, Yanhui Bian, Wenjiao Feng, Ruichen Zhang, Yinqiu Liu, Jiacheng Wang, Gang Sun, Dusit (Tao) Niyato, Hongfang Yu, Abbas Jamalipour, and Siwen Mao. Ai reasoning for wireless communications and networking: A survey and perspectives. *ACM Comput. Surv.*, April 2026. Just Accepted.

[O'shea and West, 2016] Timothy J O'shea and Nathan West. Radio machine learning dataset generation with gnu radio. In *Proceedings of the GNU radio conference*, volume 1, 2016.

[O'Shea *et al.*, 2016] Timothy J. O'Shea, Johnathan Corgan, and T. Charles Clancy. Convolutional radio modulation recognition networks. In *Engineering Applications of Neural Networks*, pages 213–226, 2016.

[Qiu *et al.*, 2024] Xiangfei Qiu, Jilin Hu, Lekui Zhou, Xingjian Wu, Junyang Du, Buang Zhang, Chenjuan Guo, Aoying Zhou, Christian S. Jensen, Zhenli Sheng, and Bin Yang. TFB: Towards comprehensive and fair benchmarking of time series forecasting methods. In *Proc. VLDB Endow.*, pages 2363–2377, 2024.

[Qiu *et al.*, 2025a] Xiangfei Qiu, Xingjian Wu, Yan Lin, Chenjuan Guo, Jilin Hu, and Bin Yang. DUET: Dual clustering enhanced multivariate time series forecasting. In *SIGKDD*, pages 1185–1196, 2025.

[Qiu *et al.*, 2025b] Xiangfei Qiu, Yuhan Zhu, Zhengyu Li, Xingjian Wu, Bin Yang, and Jilin Hu. DAG: A dual correlation network for time series forecasting with exogenous variables. *arXiv preprint arXiv:2509.14933*, 2025.

[Qu *et al.*, 2024] Yunpeng Qu, Zhilin Lu, Rui Zeng, Jintao Wang, and Jian Wang. Enhancing automatic modulation recognition through robust global feature extraction. *IEEE Transactions on Vehicular Technology*, 74(3):4192–4207, 2024.

[Ren *et al.*, 2024] Yazhou Ren, Jingyu Pu, Chenhang Cui, Yan Zheng, Xinyue Chen, Xiaorong Pu, and Lifang He. Dynamic weighted graph fusion for deep multi-view clustering. In *Proceedings of the 33rd International Joint Conference on Artificial Intelligence*, 2024.

[Rostami *et al.*, 2025a] Mohammad Rostami, Atik Faysal, Reihaneh Gh. Roshan, Huaxia Wang, Nikhil Muralidhar, and Yu-Dong Yao. DiSC-AMC: Token- and parameter-efficient discretized statistics in-context automatic modulation classification, 2025.

[Rostami *et al.*, 2025b] Mohammad Rostami, Atik Faysal, Reihaneh Gh. Roshan, Huaxia Wang, Nikhil Muralidhar, and Yu-Dong Yao. Plug-and-Play AMC: Context is king in training-free, open-set modulation with llms. In *2025 IEEE 34th Wireless and Optical Communications Conference (WOCC)*, pages 345–350, 2025.

[Sathyanarayanan *et al.*, 2023] Venkatesh Sathyanarayanan, Peter Gerstoft, and Aly El Gamal. RML22: Realistic dataset generation for wireless modulation classification. *IEEE Transactions on Wireless Communications*, 22(11):7663–7675, 2023.

[Shao *et al.*, 2025] Mingyuan Shao, Dingzhao Li, Shaohua Hong, Jie Qi, and Haixin Sun. IQFormer: A novel transformer-based model with multi-modality fusion for automatic modulation recognition. *IEEE Transactions on Cognitive Communications and Networking*, 11(3):1623–1634, 2025.

[Wang *et al.*, 2024] Shuang Wang, Hantong Xing, Chenxu Wang, Huaji Zhou, Biao Hou, and Licheng Jiao. SigDA: A superimposed domain adaptation framework for automatic modulation classification. *IEEE Transactions on Wireless Communications*, 23(10):13159–13172, 2024.

[Wang *et al.*, 2025] Shijie Wang, Qian Guo, Lu Chen, Liang Du, Zikun Jin, Zhian Yuan, and Xinyan Liang. An association-based fusion method for speech enhancement. In *Proceedings of the 34th International Joint Conference on Artificial Intelligence*, pages 6406–6414. International Joint Conferences on Artificial Intelligence Organization, 2025.

[Wei *et al.*, 2022] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed H. Chi, Quoc V. Le, and Denny Zhou. Chain-of-thought prompting elicits reasoning in large language models. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22, Red Hook, NY, USA, 2022. Curran Associates Inc.

[Wen *et al.*, 2026] Jie Wen, Jiang Long, Xiaohuan Lu, Chengliang Liu, Xiaozhao Fang, and Yong Xu. Partial multiview incomplete multilabel learning via uncertainty-driven reliable dynamic fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48(1):236–250, 2026.

[Wu *et al.*, 2025] Mei Wu, Yiqian Lin, Tianfan Jiang, and Wenchao Weng. Mhgnet: Multi-heterogeneous graph neural network for traffic prediction. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5, 2025.

[Yang *et al.*, 2025] Han Yang, Ke Liu, Wenchao Weng, Weichen Dai, Andrzej Cichocki, and Wanzeng Kong. MPF-DAN: Multi-perspective feature dynamic adaptation network for domain adaptive object detection. In *2025 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2025.

[Zhang *et al.*, 2023a] Jiawei Zhang, Tiantian Wang, Zhixi Feng, and Shuyuan Yang. Amc-net: An effective network for automatic modulation classification. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5, 2023.

[Zhang *et al.*, 2023b] Jiawei Zhang, Tiantian Wang, Zhixi Feng, and Shuyuan Yang. Toward the automatic modulation classification with adaptive wavelet network. *IEEE Transactions on Cognitive Communications and Networking*, 9(3):549–563, 2023.

[Zhang *et al.*, 2023c] Leying Zhang, Xiaokang Deng, and Yu Lu. Segment anything model (sam) for medical image segmentation: A preliminary review. In *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 4187–4194, 2023.

[Zhao *et al.*, 2025] Yurui Zhao, Xiang Wang, Shuya Cao, and Zhitao Huang. Zero-shot automatic modulation recognition using a large vision-language model. *IEEE Transactions on Communications*, 73(12):15765–15782, 2025.