In-Context Automatic Modulation Classification: First Steps Towards One LLM to Predict Them All

1[nd] Mohammad Rostami
Electrical and Computer Engineering
Rowan University
Glassboro, NJ
rostami23@students.rowan.edu

2[st] Atik Faysal
Electrical and Computer Engineering
Rowan University
Glassboro, NJ
faysal24@rowan.edu

3[rd] Reihaneh Gh. Roshan
Computer Science
Stevens Institute of Technology
Hoboken, NJ
rghasemi@stevens.edu

4[th] Huaxia Wang
Electrical and Computer Engineering
Rowan University
Glassboro, NJ
wanghu@rowan.edu

5[th] Nikhil Muralidhar
Computer Science
Stevens Institute of Technology
Hoboken, NJ
nmurali1@stevens.edu

***Abstract*—Automatic modulation classification (AMC) is critical for efficient spectrum management and robust wireless communications. However, AMC remains challenging due to the complex interplay of signal interference and noise. In this work, we propose an innovative framework that integrates traditional signal processing techniques with large language models (LLMs) to address AMC. Our approach leverages higher-order statistics and cumulant estimation to convert quantitative signal features into structured natural language prompts. By incorporating exemplar contexts into these prompts, our method exploits the LLM's inherent familiarity with classical signal processing, enabling effective one-shot classification without additional training or preprocessing (e.g., denoising). Experimental evaluations on synthetically generated datasets—spanning both noiseless and noisy conditions—demonstrate that our framework achieves competitive performance across diverse modulation schemes and signal-to-noise ratios. Moreover, our approach paves the way for foundation models in wireless communications that are robust across varying channel conditions, significantly reducing the expense associated with developing channel-specific models. Overall, this work lays the foundation for scalable, interpretable, and versatile signal classification systems in next-generation wireless networks.**

***Index Terms*—large language models, modulation classification, transformer, classification**

I. INTRODUCTION

AUTOMATIC modulation classification (AMC) is a critical technology in modern wireless communications, underpinning applications in cognitive radio, spectrum sensing, and interference management. By accurately identifying modulation schemes, AMC facilitates efficient spectrum utilization and enhances the adaptability and reliability of communication networks. However, despite its significance, AMC remains a challenging problem due to the complex interplay of signals with ambient noise and other channel impairments.

Early approaches to AMC predominantly relied on traditional machine learning techniques and handcrafted feature extraction, but these methods have gradually been superseded by deep learning models. For instance, convolutional neural networks (CNNs) have been extensively applied in this domain. Peng et al. [1] transformed raw modulated signals into constellation diagrams to serve as inputs for CNN architectures such as AlexNet, while subsequent studies introduced variants like MCNet [2] that leveraged multiple convolutional blocks with skip connections and asymmetric kernels to capture spatio-temporal correlations. In another investigation, a CNN framework was explored using multiple signal representations—constellation diagrams, ambiguity functions, and eye diagrams—with empirical results indicating that combining these representations could improve classification accuracy [3].

More recently, attention-based models, particularly transformers, have garnered significant interest for AMC. Transformers, introduced by Vaswani et al. [4], are adept at processing sequential data by dynamically focusing on the most relevant parts of the input. Cai et al. [5] demonstrated that a transformer network (TRN)-based approach can capture long-range dependencies more effectively than conventional CNNs, which is especially beneficial in low signal-to-noise ratio (SNR) scenarios. Building on this idea, Kong et al. [6] further improved transformer-based AMC by incorporating convolutional embedding and attention pooling techniques, and Qu et al. [7] combined transformers with long short-term memory (LSTM) networks to address long-range temporal dependencies and enhance robustness against noise through novel data augmentation strategies.

Despite these advancements, many existing methods are developed under idealized, noise-free conditions [8], [9]. Such conditions are rarely encountered in real-world environments where signals are typically contaminated with noise. Moreover, several approaches struggle with fine-tuning downstream tasks when only limited labeled data is available [10], [11].

cite DenoMae, DenoMaev2, nmtransformer?

Addressing these limitations, recent work has begun to explore more sophisticated frameworks that can balance performance across varying SNR conditions. Gao et al. in the MOE-AMC study [12] proposed a Mixture-of-Experts (MoE) approach that integrates a ResNet-based model for high SNR signals with a transformer-based model for low SNR signals, achieving a more balanced and robust classification performance. Complementing this, the review by Jafarigol et al. [13] provided a comprehensive analysis of contemporary AI/ML-based AMC models, highlighting both recent advances and persistent challenges—particularly the difficulty of maintaining accuracy in noisy environments and the need for efficient models that operate under constrained computational resources.

In this work, we take a novel direction by leveraging a pretrained large language model (LLM)-based approach for automatic modulation classification (AMC). Our method utilizes traditional modulation statistics and cumulants, capitalizing on the LLM's inherent familiarity with classical signal processing techniques. This capability enables our framework to perform competitive one-shot modulation classification without the need for exhaustive training across all SNRs, modulation types, or channel conditions—and without the computational overhead of extra preprocessing steps such as denoising. By transforming quantitative signal features into natural language descriptions, our approach bridges the gap between classical signal processing and modern AI, achieving robust performance even in challenging, noisy scenarios.

Our contributions include:

- A novel in-context approach for AMC using LLMs that simplifies the processing pipeline.
- A one-shot classification method that eliminates the need for extensive training data.
- Competitive performance under both noiseless and noisy conditions.
- New possibilities for integrating multimodal information in future wireless communication systems.

Beyond improving AMC performance, our approach paves the way for developing foundation models for wireless communications. Such foundation models can provide a unified solution that is robust across diverse scenarios, reducing the need for channel-specific models and lowering the associated development and deployment costs. By enabling models that generalize well without extensive retraining, our method has the potential to streamline the design of wireless communication systems, making them more adaptable, scalable, and cost-effective.

Overall, this work not only simplifies modulation classification but also sets the stage for further research in robust, scalable, and versatile signal processing using large language models.

The remainder of the paper is organized as follows. Section II presents the theoretical background, with an emphasis on high-order cumulants and their application in feature extraction. Section III details our proposed method and the process of prompt creation for our LLM-based approach. In Section IV, we report experimental results obtained on multiple datasets, and Section V concludes the paper with a discussion of future research directions.

II. Theoretical Background

check for palagrism with https://www.mdpi.com/1424-8220/24/2/701

In this section, we introduce the theoretical foundations underlying our work. We start by reviewing the key concepts of higher-order statistics and cumulants, which are essential for capturing the intrinsic features of signals. The following section explains how our framework leverages these statistical measures to construct effective prompts for our LLM-based approach.

A. High-Order Statistics

Statistical moments offer a quantitative description of the shape and characteristics of a random variable's probability distribution. In particular, the nth-order moment of a random variable x is defined as

$$m_n(x) = \int x^n f(x)\, dx, \tag{1}$$

where $f(x)$ represents the probability density function of x. The existence of the nth-order moment guarantees that all moments of lower order are also defined. We refer to $m_n(x - \mu)$ as the *central moment*, where μ is the mean, and $m_n(x/\sigma)$ as the *normalized moment*, with σ denoting the standard deviation.

Cumulants, denoted by c_n, provide an alternative distribution characterization. While the first three cumulants correspond to the mean, variance, and skewness, respectively, cumulants of order four and above capture additional aspects of the distribution's shape that are not directly reflected in the moments. The cumulant-generating function is given by

$$C_t(x) = \log\left(E\left[e^{tx}\right]\right), \tag{2}$$

which can be expanded into a power series as

$$C_t(x) = \sum_{n=1}^{\infty} \frac{c_n t^n}{n!}. \tag{3}$$

By differentiating $C_t(x)$ n times and evaluating at $t = 0$, one can obtain the nth cumulant. A noteworthy property of cumulants is their additive behavior for independent random variables:

$$c_n(X+Y) = c_n(X) + c_n(Y). \tag{4}$$

In practical applications, cumulants are often computed indirectly from moments. For example, the following relationships are commonly used [14]:

$$c_{4,0} = m_{4,0} - 3m_{2,0}^2, \tag{5}$$
$$c_{4,1} = m_{4,1} - 3m_{2,1}m_{2,0}, \tag{6}$$
$$c_{4,2} = m_{4,2} - |m_{2,0}|^2 - 2m_{2,1}^2, \tag{7}$$
$$c_{6,0} = m_{6,0} - 15m_{4,0}m_{2,0} + 30m_{3,0}^2, \tag{8}$$
$$c_{6,3} = m_{6,3} - 9c_{4,2}m_{2,1} - 6m_{2,1}^2, \tag{9}$$
$$c_{8,0} = m_{8,0} - 28m_{6,0}m_{2,0} - 35m_{4,2}^2 + 420m_{4,0}m_{2,0}^2 - 630m_{4,0}^2. \tag{10}$$

Here, $m_{q,p}$ represents the mixed moment of the signal, calculated as

$$m_{q,p} = \frac{1}{N}\sum_{n=1}^{N} x[n]^q\, x[n]^{p-q}, \tag{11}$$

with N being the total number of samples. This set of formulas facilitates the efficient computation of high-order cumulants, which are subsequently used as features in our classification algorithm.

III. Method

In this section, we present our novel approach for modulation classification, which leverages higher-order statistics and cumulant estimation via prompt-based text generation using a pretrained language model. Our method consists of three key stages: (i) Signal Summarization, (ii) Example Context Construction, and (iii) Prompt Formulation. By converting quantitative signal features into natural language descriptions, our framework not only standardizes feature representations across diverse scenarios but also integrates auxiliary observations to provide robust contextual cues for the classification task.

Specifically, for each data sample represented by the complex-valued signal x, we compute a comprehensive statistical summary. This summary is derived from global descriptors (e.g., the number of observations, minimum, maximum, mean, variance, skewness, and kurtosis) and from higher-order moments computed over both the real and imaginary components. In addition, we calculate the kth-order k-statistic, which serves as an unbiased estimator of the kth cumulant. To transform this structured numerical summary into text, we adopt a linearization technique [15] that maps the K-dimensional feature vector into a sequential format:

$$\text{linearize}(x) = \{[c_k : x_k]\}_{k=1}^{K}, \tag{12}$$

where c_k denotes the name of the kth feature and x_k its corresponding value.

The final prompt is constructed by concatenating four elements: an instruction block (I) that specifies the task, context (C) derived from exemplar signals, the linearized statistical summary, and a directive suffix (S) that instructs the model to output a single modulation type from a predefined list (e.g., 4ASK, 4PAM, 8ASK, 16PAM, CPFSK, DQPSK, GFSK, GMSK, OQPSK, OOK). Formally, the complete prompt is given by:

$$d = \text{LLM}\big(I, C,\ \text{linearize}(x),\ S\big). \tag{13}$$

This resulting textual description d succinctly encapsulates the quantitative behavior of the signal and provides clear guidance to the LLM.

A. Signal Summarization

We compute key descriptive statistics and cumulant-related metrics to obtain a robust statistical characterization of each signal. These statistical summaries serve as the foundation for both individual signal queries and for constructing exemplar contexts—reference points with known modulation labels that help the model infer the underlying patterns associated with each modulation type. Algorithm 1 details the procedure used to convert a raw signal vector x into a succinct string d that encapsulates its statistical properties.

B. Example Context Construction

To guide the language model effectively, we maintain a set of exemplar signals—one for each modulation type. For each exemplar, the same statistical summarization is performed using our `get_stats_context` function. The resulting exemplar summaries, together with their known modulation labels, are concatenated into a set of formatted examples that serve as reference points during classification. These examples

Algorithm 1: Signal Summarization

Input: x (Vector of signal values)
Output: d (A string summarizing cumulants and descriptive statistics)

```
stats ← {
  "nobs": x.nobs,
  "min": x.min,
  "max": x.max,
  "mean": x.mean,
  "variance": x.variance,
  "skewness": x.skewness,
  "kurtosis": x.kurtosis
};
for i ← 0 to 7 do
    stats["moment_i"] ← moment(signal,
     order = i);
for i ← 1 to 4 do
    stats["kstat_i"] ← kstat(signal, i);
for i ← 1 to 2 do
    stats["kstatvar_i"] ← kstatvar(signal,
     i);
d ← dict_to_string(stats);
return d;
```

provide essential contextual cues that help the model understand the statistical characteristics defining each modulation type.

These exemplars are crucial because they provide concrete contextual cues regarding the statistical characteristics that define each modulation type. By including these examples in the prompt, the language model gains a clear understanding of the expected patterns and distributions inherent to each class. This not only helps in disambiguating similar signal features across different modulation schemes but also enhances the model's ability to generalize to new, unseen signals—especially in scenarios where noise or other channel impairments might obscure the raw statistical information.

C. Prompt Formulation

The final prompt is assembled by combining three components:

- **Instruction Block** (I): A concise set of instructions that mandates the selection of a modulation type from a predefined list (e.g., "4ASK," "4PAM," "8ASK," etc.) and emphasizes that the output must consist of a single valid answer.
- **Exemplar Summaries** (C): The formatted examples derived from the exemplar signals, which display both the overall statistical summaries and their corresponding ground-truth modulation types.
- **Query Section** (linearize$(x) + S$): The statistical summary for the test signal, concatenated with a directive suffix (S) that reinforces the task.

This carefully designed prompt minimizes ambiguity and ensures that the language model's output adheres strictly to the allowed modulation types.

I think we need to add a prompt example.

Add a figure

D. Generation and Post-Processing

Once the prompt is constructed, it is tokenized and submitted to our fine-tuned causal language model. To encourage deterministic responses, the model is configured with constrained generation parameters (e.g., a low temperature and a restricted maximum token count). After the model generates its output, post-processing is performed to remove extraneous characters and extract the predicted modulation type by matching the response against the predefined list of valid classes. Finally, the predicted modulation type is compared with the ground truth label (inferred from the file name) to compute an accuracy metric.

IV. Experiments and Results

What the help are noisy and noiseless conditions? better name? you can leave comments with green todo

We evaluate our context-aware automatic modulation classification framework on a synthetically generated dataset. The dataset consists of signals produced under two conditions: *noiseless* and *noisy*. The generated signals span ten modulation formats (4ASK, 4PAM, 8ASK, 16PAM, CPFSK, DQPSK, GFSK, GMSK, OOK, OQPSK) with SNR values varied over a defined range to simulate different channel conditions. Our evaluation focuses on overall classification accuracy, robustness under both noisy and noiseless conditions, and the impact of different prompt components.

A. Dataset Generation

Our testing dataset is generated to include 4,000 signals, evenly split between *noiseless* and *noisy* conditions. For the downstream classification task, 200 samples are used for testing, evenly distributed among the 10 modulation classes, with SNR values ranging from -10 dB to 10 dB. This setup allows us to assess the performance of our framework under ideal conditions as well as in the presence of channel-induced noise.

TABLE I: Modulation Classification Performance Across Models, Prompt Contexts, and Model Sizes

Model	Parameters	Prompt	Signal	Accuracy (%)	Cleaned Accuracy (%)
DeepSeek-R1-Distill-Qwen	7B	$I + S$	Noiseless	12.6	14.35
DeepSeek-R1-Distill-Qwen	7B	$I + S$	Noisy	11.8	13.66
DeepSeek-R1-Distill-Qwen	7B	$I + C + S$	Noiseless	19.3	54.52
DeepSeek-R1-Distill-Qwen	7B	$I + C + S$	Noisy	5.2	27.81
DeepSeek-R1-Distill-Qwen	32B	$I + S$	Noiseless	9.1	10.83
DeepSeek-R1-Distill-Qwen	32B	$I + S$	Noisy	7.9	9.58
DeepSeek-R1-Distill-Qwen	32B	$I + C + S$	Noiseless	**58.8**	**61.44**
DeepSeek-R1-Distill-Qwen	32B	$I + C + S$	Noisy	**47.8**	**53.52**
OpenAI-o3-mini	200B	$I + C + S$	Noisy	**69.92**	**72.4**

$I + S$ refers to the scenario where the language model receives only a concise block of instructions to select the modulation type from a predefined list, without any additional exemplar context. In contrast, *Example* $I + C + S$ incorporates exemplar summaries derived from known modulation signals into the prompt. The terms *Noiseless* simulate ideal channel conditions with minimal interference, while *Noisy* signals incorporate ambient or channel-induced noise. *Cleaned Accuracy* refers to the accuracy obtained after removing responses that did not identify any modulations from the predictions.

B. Experimental Setup

For each data sample, the raw complex-valued signal $x_{i,t}$ is processed to extract its I/Q components, and a comprehensive statistical summary is computed as described in Section III. This summary is then transformed into a natural language prompt using our linearization technique. The prompt, which includes a instructions I, context-setting prefix C, the linearized statistical summary, and a directive suffix S, is fed into our fine-tuned language model configured with deterministic decoding parameters. The model's output, restricted to a predefined list of modulation types, is compared against the ground truth to compute classification accuracy.

C. Overall Classification Performance

Table I summarizes the overall classification accuracy of our proposed method, comparing performance across different prompt contexts, signal conditions, and model sizes. Our experiments reveal several key insights regarding the influence of exemplar context and model capacity on classification performance.

First, incorporating exemplar context into the prompt (i.e., using the $I + C + S$ configuration) substantially improves performance compared to using only the instruction and suffix ($I + S$). For example, with the DeepSeek-R1-Distill-Qwen-7B model, the inclusion of exemplar context increases the cleaned accuracy from 14.35% under noiseless conditions to 54.52%, while under noisy conditions accuracy improves from 13.66% to 27.81%.

Second, the effect of model size is evident. The DeepSeek-R1-Distill-Qwen-32B model, which has significantly more parameters than the 7B variant, achieves a cleaned accuracy of 61.44% (noiseless) and 53.52% (noisy) with the $I + C + S$ prompt—an improvement of over 40% relative to its $I + S$ configuration. This indicates that larger models, with their enhanced representational capacity, can better exploit the rich contextual information provided by exemplar summaries.

Finally, the o3-mini model, boasting 200 billion parameters, attains the highest performance, particularly under noisy conditions, with an overall accuracy of 69.92% (72.4% cleaned). This suggests that scaling up the model further yields substantial gains, even when explicit channel or SNR information is not provided beyond what is captured in the signal summaries.

Moreover, our results show that the performance on noiseless signals is generally better than on noisy signals, implying that while the LLM can recognize modulation types in the presence of noise, its performance degrades as noise increases. This observation indicates that incorporating additional context about SNR may further improve classification accuracy.

V. Conclusion and Future Work

In this paper, we presented a novel in-context approach for automatic modulation classification that leverages higher-order statistics and cumulant estimation via prompt-based classification using a pretrained large language model. By transforming quantitative signal features into natural language descriptions, our framework effectively captures the intrinsic characteristics of complex-valued signals across diverse scenarios without relying on explicit channel or SNR inputs. Experimental results on synthetically generated datasets demonstrate that incorporating exemplar context into the prompt significantly boosts classification accuracy, and that larger models—such as the 200-billion-parameter o3-mini—can further enhance performance, particularly under noisy conditions.

Future Work. In future work, we plan to explore several avenues for improvement and expansion:

1) **Enhanced Prompt Engineering:** Investigate additional prompt structures and incorporate explicit channel and SNR information into the prompt to further refine the model's performance.
2) **Model Scaling and Fine-Tuning:** Evaluate the performance of even larger language models and develop specialized fine-tuning strategies tailored specifically for modulation classification tasks.
3) **Model Robustness:** Address the inherent stochasticity introduced by non-zero generation temperature, which may yield different outputs for identical inputs. We aim to develop alternative approaches to increase robustness, particularly for SNR conditions that the model has not encountered during training.
4) **Fine-Tuning on Generated Data:** Fine-tune the model on our generated dataset to better capture the nuances of the synthesized signals and further improve classification performance.
5) **Reasoning Fine-Tuning:** Leverage the model's capability to generate explanations for its decisions by using these generated rationales as additional supervisory signals during fine-tuning.
6) **Explainability:** Investigate methods to enhance the interpretability of the model's predictions and to extract meaningful insights from the generated textual descriptions.
7) **Real-World Signal Validation:** Extend our evaluation to real-world datasets to assess the robustness and generalizability of our approach in operational environments, including the integration of adaptive noise mitigation techniques.
8) **Multimodal Integration:** Explore incorporating additional modalities, such as constellation diagrams (images), to complement high-order statistical summaries. This multimodal approach could mitigate the information loss inherent in summarizing signals solely via cumulants, and may pave the way for leveraging vision-language models (VLMs) and advanced self-supervised fine-tuning methods.
9) **Broader Application Domains:** Examine the applicability of our in-context methodology to related domains, such as spectrum sensing and cognitive radio, with the long-term goal of developing a foundation model for wireless communications.

Overall, our work represents a significant first step toward leveraging large language models for automatic modulation classification. We are confident that further advancements in prompt design, model architecture, and domain-specific integration will unlock new possibilities for robust and efficient signal processing in wireless communications.

REFERENCES

[1] S. Peng, H. Jiang, H. Wang, H. Alwageed, and Y.-D. Yao, "Modulation classification using convolutional neural network based deep learning model," in *2017 26th Wireless and Optical Communication Conference (WOCC)*, pp. 1–5, IEEE, 2017.

[2] T. Huynh-The, C.-H. Hua, Q.-V. Pham, and D.-S. Kim, "Mcnet: An efficient cnn architecture for robust automatic modulation classification," *IEEE Communications Letters*, vol. 24, no. 4, pp. 811–815, 2020.

[3] A. Samarkandi, A. Almarhabi, H. Alhazmi, and Y.-D. Yao, "Combined signal representations for modulation classification using deep learning: Ambiguity function, constellation diagram, and eye diagram," in *2023 32nd Wireless and Optical Communications Conference (WOCC)*, pp. 1–4, 2023.

[4] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[5] J. Cai, F. Gan, X. Cao, and W. Liu, "Signal modulation classification based on the transformer network," *IEEE Transactions on Cognitive Communications and Networking*, vol. 8, no. 3, pp. 1348–1357, 2022.

[6] W. Kong, X. Jiao, Y. Xu, B. Zhang, and Q. Yang, "A transformer-based contrastive semi-supervised learning framework for automatic modulation recognition," *IEEE Transactions on Cognitive Communications and Networking*, 2023.

[7] Y. Qu, Z. Lu, R. Zeng, J. Wang, and J. Wang, "Enhancing automatic modulation recognition through robust global feature extraction," *arXiv preprint arXiv:2401.01056*, 2024.

[8] S. Peng, H. Jiang, H. Wang, H. Alwageed, Y. Zhou, M. M. Sebdani, and Y.-D. Yao, "Modulation classification based on signal constellation diagrams and deep learning," *IEEE transactions on neural networks and learning systems*, vol. 30, no. 3, pp. 718–727, 2018.

[9] F. Meng, P. Chen, L. Wu, and X. Wang, "Automatic modulation classification: A deep learning enabled approach," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 11, pp. 10760–10772, 2018.

[10] D. Hong, Z. Zhang, and X. Xu, "Automatic modulation classification using recurrent neural networks," in *2017 3rd IEEE International Conference on Computer and Communications (ICCC)*, pp. 695–700, IEEE, 2017.

[11] S. Ramjee, S. Ju, D. Yang, X. Liu, A. E. Gamal, and Y. C. Eldar, "Fast deep learning for automatic modulation classification," *arXiv preprint arXiv:1901.05850*, 2019.

[12] J. Gao, Q. Cao, and Y. Chen, "Moe-amc: Enhancing automatic modulation classification performance using mixture-of-experts," *arXiv preprint arXiv:2312.02298*, 2023.

[13] E. Jafarigol, B. Alaghband, A. Gilanpour, S. Hosseinipoor, and M. Mirmozafari, "Ai/ml-based automatic modulation recognition: Recent trends and future possibilities," *arXiv preprint arXiv:2502.05315*, 2025.

[14] E. Azzouz and A. Nandi, "Automatic identification of digital modulation types," *Signal Processing*, vol. 47, no. 1, pp. 55–69, 1995.

[15] Z. Wang, C. Gao, C. Xiao, and J. Sun, "Meditab: scaling medical tabular data predictors via data consolidation, enrichment, and refinement," *arXiv preprint arXiv:2305.12081*, 2023.