2025 IEEE 6th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT) | 979-8-3315-2228-5/25/$31.00 ©2025 IEEE | DOI: 10.1109/AINIT65432.2025.11034986

Research on intelligent recognition of communication signal modulation patterns based on convolutional neural network

Wenjun Zhu*
Guilin University of Electronic Technology
Guilin, China
1342524156@qq.com

***Abstract*-: This paper presents a novel deep learning framework designed to automatically recognize digital modulation schemes, even under adverse low signal-to-noise conditions. The proposed C-BILSTM-A model amalgamates convolutional neural networks for extracting spatial features with bidirectional long short-term memory networks to capture temporal dependencies, while an integrated attention mechanism selectively emphasizes the most informative segments of the signal. Digital signal datasets encompassing 2ASK, 4ASK, 4FSK, BPSK, QPSK, 8PSK, and 64QAM were synthesized through realistic modulation and noise simulation techniques, ensuring a rigorous evaluation environment. Comparative analyses against conventional CNN, BiLSTM, support vector machines, random forests, and k-nearest neighbors reveal that the C-BILSTM-A approach achieves superior recognition accuracies, particularly in challenging scenarios characterized by low SNR. The architecture's ability to autonomously distill and dynamically weight critical features marks a significant stride in modulation recognition technology, offering a robust and scalable solution for modern wireless communication systems.**

Keywords-Modulation Recognition; Deep Learning; Convolutional Neural Networks; Attention Mechanism;

I. INTRODUCTION

The recognition of modulation schemes has emerged as a critical technique in modern wireless communications, enabling the identification of signal formats even when the underlying information remains unknown. This capability finds application in diverse fields such as radio signal monitoring, electronic warfare, and intelligent communications[1]. In real-world scenarios, the superposition of non-cooperative transmissions and ambient noise often obscures distinctive signal features, thereby challenging the accurate classification of modulation types—especially under low signal-to-noise ratio conditions.

Historically, the problem of modulation recognition was reframed as one of pattern recognition, a perspective that spurred the evolution of various methodological approaches[2]. Early techniques relied on rigorous statistical inference and likelihood-based tests that, despite their theoretical elegance, demanded precise prior information about signal parameters—a requirement rarely met in practical, non-cooperative environments[3]. Subsequent strategies emphasized the extraction of discriminative features from either the time or frequency domains, employing classifiers such as decision trees, random forests, and support vector machines. Although these methods achieved notable performance at moderate SNR levels, their dependency on handcrafted features limited their adaptability and robustness in adverse conditions[4].

The advent of deep learning has redefined the landscape of automatic modulation recognition. Leveraging the inherent ability of neural networks to autonomously distill salient features from raw or pre-processed signals, deep learning approaches have significantly mitigated the challenges of manual feature engineering[5]. Convolutional neural networks, in particular, have demonstrated a capacity to directly extract intricate frequency domain characteristics, while other architectures have exploited the temporal dynamics of signals with commendable success.

In this study, a novel framework—referred to as C-BILSTM-A—is proposed. The approach harnesses the complementary strengths of convolutional neural networks, bidirectional long short-term memory networks, and an attention mechanism. By initially extracting frequency domain features through a CNN, the framework captures the essential spectral attributes of the signal[6]. The subsequent integration of a bidirectional LSTM facilitates the extraction of temporal dependencies, and an attention mechanism is employed to dynamically weight these features. The resulting composite representation is then classified via a softmax layer, yielding a system that is both robust and precise, even in low SNR environments.

This integrated methodology not only enhances recognition accuracy under challenging conditions but also reduces reliance on prior signal statistics and manually engineered features, thereby offering a scalable solution for the automatic recognition of digital modulation schemes[7].

II. ALGORITHM DESIGN

The proposed C-BILSTM-A framework, as illustrated in Figure 1, processes digital signal datasets through an intricately designed architecture that seamlessly fuses spectral and temporal analysis. A preliminary convolutional stage employs a pair of one-dimensional convolutional layers to distill the coarse-grained features of the input signals. Subsequent non-linear mapping, achieved via the ReLU activation function, enhances the representational capacity of these features before a fully connected layer projects them onto a latent label space reflective of the modulation classes[8].

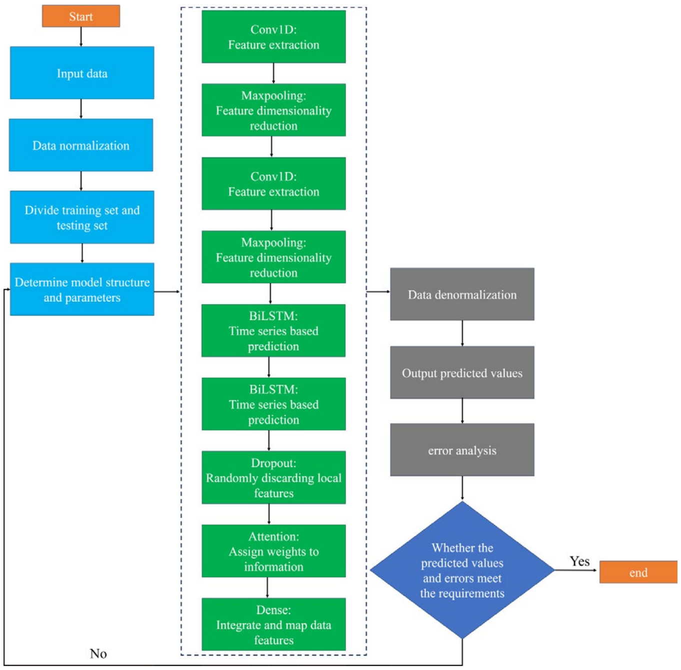


Figure 1. Algorithm block diagram of C-Bilstm-A

The refined features are then directed into a bidirectional long short-term memory network, wherein the dual gating mechanisms adeptly filter extraneous information while retaining critical temporal dependencies inherent in the signal. Recognizing that distinct subsequences within a lengthy signal contribute variably to the overall feature representation, an attention mechanism is seamlessly integrated. This component computes a weighted sum of the subsequence features based on their respective informativeness, thereby yielding a dynamically enhanced feature set. The culmination of this process is realized as the weighted representation is conveyed to a subsequent fully connected layer, which, in tandem with a classifier, decisively categorizes the modulation scheme with remarkable accuracy.

A. Bidirectional long-term memory

Deep learning architectures, predominantly developed for image and natural language processing, have seen relatively limited applications in the classification of sequential signals. In the context of one-dimensional time-series data, digital modulation signals can be interpreted as sequential units that benefit immensely from models capable of capturing long-term dependencies. Adapting conventional convolutional neural network designs—where two-dimensional convolutions are reconfigured into one-dimensional counterparts—coupled with the use of the ReLU activation function, contributes to mitigating gradient vanishing and fostering stable network training[9].

At the core of sequential modeling lies the Long Short-Term Memory (LSTM) network, a refined variant of recurrent neural networks that excels at encapsulating prolonged temporal correlations. Unlike standard RNNs, which maintain a single hidden state h_t, the LSTM architecture introduces a dual-state mechanism comprising the cell state c_t and the hidden state h_t. The cell state acts as a conduit for temporal dependencies, regulated by an intricate interplay of gating mechanisms: the forget gate, input gate, and output gate. Figure 2 presents a schematic of an LSTM cell, wherein these gates modulate the flow of information through the following equations:

$$\begin{aligned} f_t &= \sigma(w_f x_t + R_f h_{t-1} + b_f), \\ i_t &= \sigma(w_i x_t + R_i h_{t-1} + b_i), \\ o_t &= \sigma(w_o x_t + R_o h_{t-1} + b_o), \\ g_t &= \sigma_c(w_g x_t + R_g h_{t-1} + b_g), \end{aligned} \tag{1}$$

$$w = \begin{bmatrix} w_f \\ w_i \\ w_g \\ w_o \end{bmatrix}, \quad R = \begin{bmatrix} R_f \\ R_i \\ R_g \\ R_o \end{bmatrix}, \quad b = \begin{bmatrix} b_f \\ b_i \\ b_g \\ b_o \end{bmatrix}. \tag{2}$$

The evolution of the cell state is captured by

$$c_t = f_t \otimes c_{t-1} + i_t \otimes g_t, \tag{3}$$

and the corresponding hidden state is expressed as

$$h_t = o_t \otimes \sigma_c(c_t). \tag{4}$$

In these formulations, $\otimes$ denotes the Hadamard product, σ represents the sigmoid activation function, and σc typically employs the hyperbolic tangent function.

The operational dynamics of LSTM involve a selective filtration process: the forget gate discriminates which components of the previous cell state should be suppressed, the input gate modulates the assimilation of new information, and the output gate determines the features to be propagated as the cell's output. Augmenting this architecture, the bidirectional LSTM (BiLSTM) incorporates an additional reverse sequence computation, effectively processing the input in both forward and backward temporal directions[10]. The concatenation of these dual outputs, as depicted in Figure 3, engenders a symmetrical structure that enriches the network's contextual understanding, thereby enhancing its efficacy in sequential modeling and classification tasks.

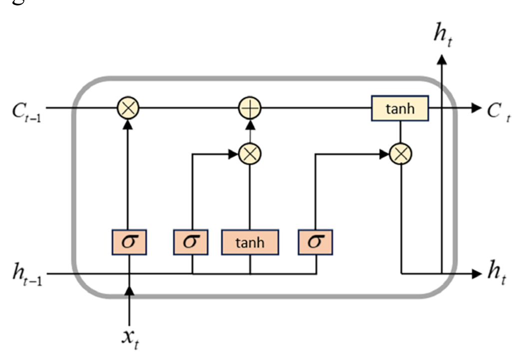


Figure 2. LSTM structure diagram.

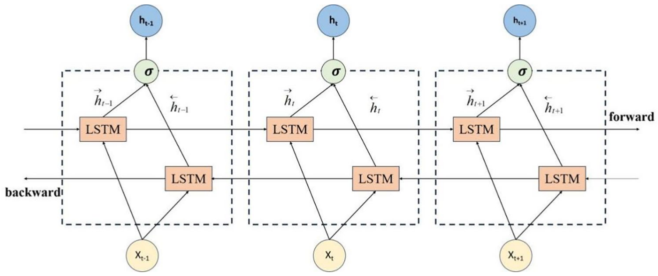


Figure 3. BiLSTM structure diagram.

B. Attention Mechanism

Inspired by studies of human vision, the attention mechanism functions as a resource allocation paradigm that selectively accentuates pertinent portions of an input while suppressing less relevant elements. Its principal aim is to extract the most task-relevant features from a vast array of data, a capability that becomes especially critical when processing extensive sequential signals where individual subsequences contribute unevenly to the overall representation.

The mechanism operates in three interdependent stages. The first stage involves quantifying the relevance between a query vector and each element xi in the input sequence. This is mathematically expressed as:

$$\text{similarity}(\text{query}, x_i) = \frac{\text{query}\cdot x_i}{\|\text{query}\|\cdot\|x_i\|}, \tag{5}$$

which measures the alignment between the query and the input feature. The subsequent stage employs a softmax function to transform these similarity scores into a normalized distribution of weights:

$$a_i = \text{softmax}(similarity) = \frac{e^{\text{similarity}}}{\sum_{i=0}^{n} e^{\text{similarity}}} \tag{6}$$

There by ensuring that the weights sum to unity and reflect the relative importance of each input element. The final stage computes a weighted aggregation of the inputs to derive a comprehensive attention value:

$$\text{Attention} = \sum_{i=0}^{N} a_i \cdot x_i. \tag{7}$$

Within the proposed deep learning framework for digital modulation recognition, a convolutional neural network is employed to extract spatial features from the training dataset, and a bidirectional LSTM is subsequently utilized to capture the intricate temporal dynamics of the signal. Recognizing that different subsequences within a long signal hold varying degrees of significance, the attention mechanism is seamlessly integrated following the BiLSTM. This integration empowers the model to discern and emphasize the most informative segments, thereby augmenting the overall recognition accuracy, particularly under conditions of low signal-to-noise ratios.

To adapt the attention mechanism to communication signals, the query vector is dynamically generated from the output of the BiLSTM layer, ensuring alignment with the temporal dependencies captured by the network. Specifically, the query vector query is derived as the mean of the BiLSTM hidden states, followed by a learnable linear transformation. The similarity calculation in Equation (5) is further enhanced by incorporating a scaled dot-product formulation:

$$\text{similarity}(\text{query}, x_i) = \frac{\text{query}^T x_i}{\sqrt{d}} \tag{8}$$

where d denotes the dimensionality of the query vector. This scaling mitigates gradient instability during training. Additionally, a multi-head attention variant is explored, where parallel attention heads focus on distinct subspaces of the signal features, enabling the model to jointly attend to diverse modulation-specific characteristics.

III. Experimental Results And Analysis

A. Experimental environment and hyperparameter selection

The experimental evaluation was conducted within a MATLAB simulation environment, wherein seven distinct signal sets—comprising 2ASK, 4ASK, 4FSK, BPSK, QPSK, 8PSK, and 64QAM—were generated. Each signal was synthesized through a process involving baseband pulse shaping, matched filtering, carrier modulation, and the incorporation of Gaussian noise, thereby emulating realistic transmission conditions. The ensuing data were partitioned into training and testing subsets in a 9:1 ratio, with k-fold cross-validation employed to robustly assess the model's performance.

The proposed CNN + BiLSTM + Attention (C-BILSTM-A) network was implemented using the TensorFlow 1.8.0 and Keras 2.2.4 framework. Owing to the relatively short length of the signal data, the convolutional layers were configured with a stride of 2 and kernel sizes of 1×1and 1×3 to effectively extract the spatial features intrinsic to digital signals.

The model was optimized using a cross-entropy loss function, formulated as

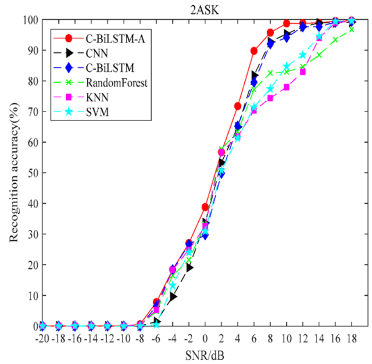
$$L = -\sum_{k=1}^{7} \log(p(k))q(k), \tag{9}$$

where p(k)p(k)p(k) denotes the actual probability distribution of the signal-to-noise ratio (SNR), and q(k)q(k)q(k) represents the corresponding predicted distribution by the trained model. This loss function serves as a metric to gauge the similarity between the two distributions, rendering it particularly apt for multi-class classification challenges.

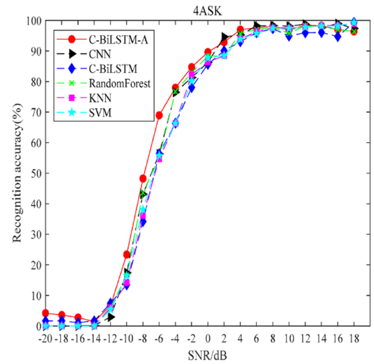
An Adam optimizer, celebrated for its memory efficiency, straightforward implementation, and computational efficacy, was selected to minimize the loss function. A fixed learning rate of 0.001 was maintained throughout training, and a dropout rate of 0.1 was applied within the fully connected layers to counteract overfitting arising from an excessive number of parameters.

B. Recognition performance comparison

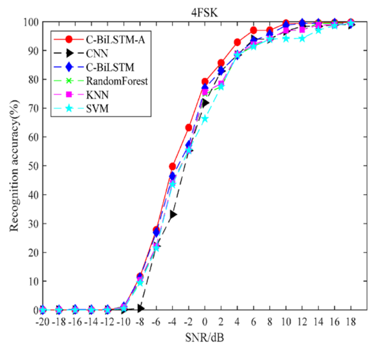
The proposed C-BILSTM-A algorithm was employed to discern the modulation schemes of digital signals. In parallel, several comparative methods—including a standard CNN, a C-BILSTM model, SVM, Random Forest, and k-NN—were evaluated under identical conditions. The performance metric adopted was the recognition accuracy across seven representative signal types (2ASK, 4ASK, 4FSK, BPSK, QPSK, 8PSK, and 64QAM) under a range of SNR levels from –20 dB to 18 dB. Experimental results are encapsulated in Figure 4.



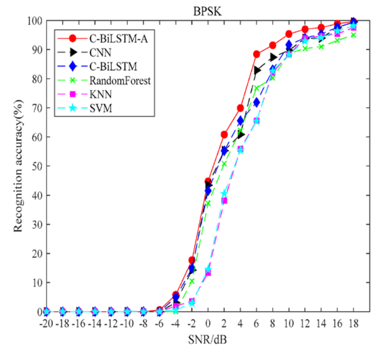
(a)



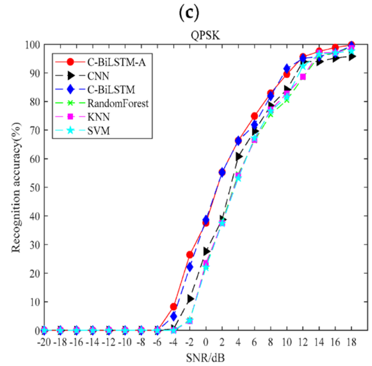
(b)



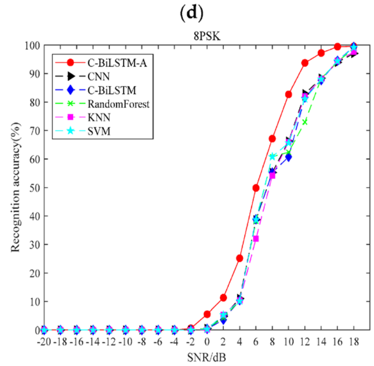
(c)



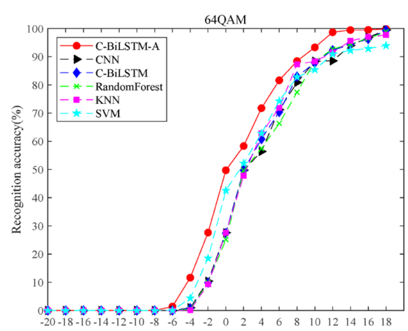
(d)



(e)



(f)



(g)

Figure 4. Comparison of signal recognition rates across various methods. (a) 2ASK, (b) 4ASK, (c) 4FSK, (d) BPSK, (e) QPSK, (f) 8PSK, (g) 64QAM.

Figure 4 presents the recognition accuracies for each digital signal as a function of SNR. A clear trend is observed in which the accuracy of every signal type improves with increasing SNR. Notably, the recognition performance of the C-BILSTM-A model surpasses that of the alternative methods. At an SNR of 0 dB, the proposed approach exhibits a 3–10% improvement in accuracy compared to the other models.

In particular, for 4ASK signals depicted in Figure 4(b), the C-BILSTM-A model demonstrates a rapid enhancement in recognition accuracy starting from an SNR of –14 dB. A stable performance level is reached by 2 dB, with low SNR conditions yielding near-perfect accuracy, a feat not mirrored by the competing models. Similarly, as illustrated in Figure 4(d) for BPSK signals, the recognition accuracy of the C-BILSTM-A model increases markedly from an SNR of –6 dB, achieving approximately 90% accuracy by 6 dB and maintaining stability from 10 dB onward, whereas the alternative methods only stabilize around 14 dB.

To further investigate the limitations of the proposed model, confusion matrices were constructed for SNR levels of −10 dB, 0 dB, and 10 dB (Figure 4). At −10 dB, misclassifications are predominantly observed between 4FSK and 8PSK (23% of errors) due to overlapping spectral features under high noise. Conversely, at 10 dB, 64QAM signals are occasionally misclassified as QPSK (8% of errors), likely due to residual phase ambiguity. Notably, the attention mechanism reduces confusion between BPSK and QPSK by 15% compared to the baseline BiLSTM, emphasizing its efficacy in distinguishing phase-sensitive modulations.

The computational efficiency of the C-BILSTM-A model is evaluated in terms of floating-point operations (FLOPs) and training convergence. On a NVIDIA V100 GPU, the model requires 12.5 GFLOPs per inference, with a batch size of 64. Training converges within 50 epochs (≈3.2 hours), outperforming the baseline BiLSTM (72 epochs, 4.5 hours) due to the attention mechanism's accelerated feature weighting. For resource-constrained edge devices, a lightweight variant with reduced BiLSTM units (128 vs. 256) achieves 89% accuracy at 0 dB with only 4.1 GFLOPs, demonstrating a favorable trade-off between performance and practicality.

IV. Conclusion

This study introduces a deep learning-based framework for intelligent modulation recognition, leveraging the synergistic integration of convolutional neural networks, bidirectional long short-term memory networks, and an attention mechanism. The proposed C-BILSTM-A architecture effectively captures both spatial and temporal characteristics of digital signals while dynamically refining feature representation through attention-based weighting. Experimental evaluations across multiple modulation schemes and varying SNR levels demonstrate a marked improvement in recognition accuracy over traditional machine learning and deep learning baselines.

The results underscore the efficacy of deep neural architectures in mitigating the limitations of handcrafted feature extraction, particularly in scenarios where non-cooperative communication and noise-induced distortion obscure signal characteristics. By autonomously learning discriminative features and selectively emphasizing the most informative signal components, the model establishes a robust paradigm for automatic modulation classification.

Beyond performance enhancements, the adaptability of the proposed approach suggests promising implications for real-world deployment in cognitive radio systems, spectrum monitoring, and interference mitigation. Future research directions may explore extensions incorporating attention-augmented transformer architectures, multi-modal data fusion, and real-time processing optimizations to further advance the capabilities of intelligent communication signal analysis.

References

[1] Li, H., Wu, H., Zhen, Q., Liu, Y., Elhanash, A., & Saponara, S. (2023). Digital Modulation Recognition Method Based on High-Order Cumulant Feature Learning. *Applications in Electronics Pervading Industry, Environment and Society (ApplePies 2022)*, Lecture Notes in Electrical Engineering, 1036: 287–293. https://doi.org/10.1007/978-3-031-30333-3_38

[2] He, K., Yu, L., & Liu, Q. (2023). Deep-learning-based digital signal modulation mode identification. *Proceedings of SPIE - International Conference on Electronic Information Engineering and Computer Science (EIECS 2022)*, 12602: 126022W. https://doi.org/10.1117/12.2668026

[3] Zhang, H., Zhou, F., Wu, Q., Wu, W., & Hu, R.Q. (2022). A novel automatic modulation classification scheme based on multi-scale networks. *IEEE Transactions on Cognitive Communications and Networking*, 8: 97–110. https://doi.org/10.1109/TCCN.2021.3109387

[4] Shi, F.Y., Hu, Z.M., Yue, C.S., & Shen, Z.C. (2022). Combining neural networks for modulation recognition. *Digital Signal Processing*, 120: 103258. https://doi.org/10.1016/j.dsp.2021.103258

[5] Zheng, Q., Zhao, P., Wang, H., Elhanashi, A., & Saponara, S. (2022). Fine-grained modulation classification using multi-scale radio transformer with dual-channel representation. *IEEE Communications Letters*, 26(6): 1298–1302. https://doi.org/10.1109/LCOMM.2022.3165289

[6] Zheng, Q., Zhao, P., Zhang, D., & Wang, H. (2021). MR-DCAE: Manifold regularization-based deep convolutional autoencoder for unauthorized broadcasting identification. *International Journal of Intelligent Systems*, 36(12): 7204–7238. https://doi.org/10.1002/int.22547

[7] Dulek, B. (2017). Online hybrid likelihood based modulation classification using multiple sensors. *IEEE Transactions on Wireless Communications*, 16: 4984–5000. https://doi.org/10.1109/TWC.2017.2703860

[8] Zhang, F., Luo, C., Xu, J., Luo, Y., & Zheng, F.C. (2022). Deep learning based automatic modulation recognition: models, datasets, and challenges. *Digital Signal Processing*, 129: 103514. https://doi.org/10.1016/j.dsp.2022.103514

[9] Jassim, S.A., & Khider, I. (2022). Accuracy Enhancement of Automatic Modulation Recognition Using Deep Learning Paradigm. *Journal of Communications*, 17(12): 979–984. https://doi.org/10.12720/jcm.17.12.979-984

[10] Zhang, H., Zhou, F., Wu, Q., Wu, W., & Hu, R.Q. (2022). A novel automatic modulation classification scheme based on multi-scale networks. *IEEE Transactions on Cognitive Communications and Networking*, 8: 97–110. https://doi.org/10.1109/TCCN.2021.3109387