Proceedings of CONF-SEML 2026 Symposium: Importance of Machine Learning Methods and Analysis in Engineering
DOI: 10.54254/2755-2721/2026.AS32102

Comparative Analysis of 1D-ResNet and CNN–BiLSTM for Automatic Modulation Classification under Low-to-Medium SNR Conditions

Jiahui Liang

School of Electromechanical Engineering, Guangdong University of Technology, Guangzhou, China
2965540549@qq.com

Abstract. Automatic Modulation Classification (AMC) plays a pivotal role in modern wireless communication systems, supporting applications such as cognitive radio spectrum management and secure military communications. However, traditional methods based on artificial features and conventional classifiers exhibit limited robustness under low signal-to-noise ratios (SNR) and struggle to adapt to dynamic signal characteristics, particularly in fading and non-Gaussian noise environments. Among deep learning approaches, lightweight one-dimensional Residual Networks (1D-ResNet) and CNN-BiLSTM architectures have attracted considerable attention due to their complementary strengths in structural and temporal feature learning. This paper systematically compares 1D-ResNet and CNN-BiLSTM using the RadioML 2018.01A dataset, which contains 24 modulation formats. Experiments are performed at 2 dB SNR intervals across the 0-14 dB range. The results demonstrate a clear SNR-dependent performance transition. The 1D-ResNet consistently outperforms the CNN-BiLSTM model in low-SNR conditions (0-4 dB), achieving classification accuracies of 0.5242 vs. 0.4804 at 0 dB. Conversely, the CNN-BiLSTM model surpasses the 1D-ResNet at higher SNR levels, reaching 0.8755 vs. 0.8545 at 10 dB.

Keywords: Automatic Modulation Classification, 1D-ResNet, CNN–BiLSTM, Deep Learning, RadioML 2018.01A

1. Introduction

Automatic Modulation Classification (AMC) is a fundamental technology in modern wireless communication systems, supporting critical applications such as cognitive radio (CR) spectrum management, electronic warfare, and adaptive transmission in 5G/6G networks [1,2]. Its primary objective is to identify the modulation scheme of received signals without prior information, directly impacting spectrum utilization efficiency and demodulation reliability in complex electromagnetic environments [1,2]. However, traditional AMC approaches based on handcrafted features and conventional machine learning classifiers often exhibit limited robustness in low signal-to-noise ratio (SNR) conditions and struggle to adapt to the dynamic characteristics of modern communication signals, particularly under fading and non-Gaussian noise [1,2]. Recent advances in deep learning have revolutionized AMC by enabling end-to-end feature learning directly from raw in-phase and quadrature (I/Q) samples, thereby reducing dependence on manual feature engineering. Among

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.AS32102

various deep learning architectures, lightweight one-dimensional Residual Networks (1D-ResNet) and CNN–BiLSTM models have attracted considerable attention due to their complementary modeling capabilities. The 1D-ResNet architecture employs residual connections to alleviate gradient degradation in deep networks, facilitating effective extraction of hierarchical structural features from time-domain signals, which is particularly beneficial under noisy and fading channel conditions. In contrast, CNN–BiLSTM architectures integrate convolutional layers for local feature extraction with bidirectional Long Short-Term Memory (BiLSTM) units to model long-term temporal dependencies. This makes them well-suited for capturing the time-varying characteristics of communication signals. However, existing studies have not yet reached a clear consensus regarding their relative performance under varying SNR conditions, which complicates architecture selection for real-world systems where channel quality is highly dynamic. To address this gap, this study conducts a systematic comparative analysis of 1D-ResNet and CNN-BiLSTM architectures for Automatic Modulation Classification (AMC) across a moderate SNR range of 0-14 dB. This investigation aims to enhance the robustness and adaptability of wireless communication systems operating under complex channel conditions. The findings offer an effective approach for improving modulation recognition performance in low-SNR environments, thereby potentially enhancing communication system stability and security.

2. Literature review

Deep learning-based AMC has been extensively investigated in recent years, with LSTM-based and ResNet-based architectures emerging as two dominant research directions. Several studies have demonstrated that LSTM-based architectures effectively model sequential dependencies in communication signals, particularly in noisy and time-varying environments [3-5]. For example, Ansari et al. proposed an optimized LSTM framework with adaptive temporal-spectral feature learning, achieving a classification accuracy of 99.87% at −5 dB SNR and significantly outperforming traditional machine learning methods [3]. Parallel to the progress in LSTM-based models, ResNet-based AMC methods have focused on improving deep feature extraction under challenging channel conditions by leveraging residual connections [6]. Building on this line of research, Shankar et al. proposed a ResNet-based architecture LResNet tailored for AMC, then derived its rotated binarized version RBLResNet to reduce complexity for edge deployment [7]. Beyond conventional AMC tasks, Choi et al. solved temporal information loss in ResNet-based ASV models caused by convolution stride and ASP incompatibility by proposing TB-ResNet with transposed convolution TB-ResBlocks [8]. Despite these advances, existing studies predominantly evaluate CNN-BiLSTM/LSTM- or ResNet-based models under isolated SNR conditions, often focusing on either extreme low or high SNR regimes. Thus, a unified and systematic comparison of their performance across moderate and dynamically varying SNR ranges remains insufficient, which motivates the comparative analysis conducted in this study. Moreover, this fragmented approach limits the understanding of their performance across the contiguous, moderate SNR range (0-14 dB) that is critical for many practical applications.

3. Model architecture

3.1. 1D-ResNet architecture

Residual Networks (ResNet), originally proposed by He et al., mitigate the degradation problem in deep neural networks by introducing identity shortcut connections [9]. Instead of directly learning a mapping H(x), ResNet learns a residual function:

$$y = F(x, W) + x \tag{1}$$

DOI: 10.54254/2755-2721/2026.AS32102

This formulation enables efficient gradient propagation and stable training of deep models, making ResNet well-suited for extracting hierarchical structural features from noisy communication signals.

In this study, a 1D-ResNet is designed to process complex baseband IQ samples, as illustrated in Figure 1. The input consists of two channels (I/Q) with 1024 samples. An initial 1D convolution followed by a max pooling layer is employed for early feature extraction and temporal downsampling. Subsequently, three residual block groups with increasing channel dimensions (32, 64, and 128) are constructed, each containing two residual blocks. Within each residual block, identity-based shortcut connections are introduced to enable direct information and gradient flow across layers. By learning residual mappings rather than complete feature transformations, these shortcut connections effectively alleviate the gradient vanishing and network degradation problems in deep architectures.

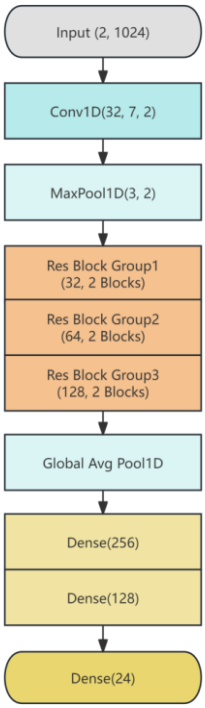

Figure 1. The architecture of 1D-ResNet

3.2. CNN–BiLSTM

Architecture Long Short-Term Memory (LSTM), introduced by Hochreiter and Schmidhuber [10], enhances recurrent neural networks by incorporating gated memory cells to model long-term temporal dependencies. The LSTM cell updates are governed by:

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \tag{2}$$

$$h_t = o_t \odot \tanh(c_t) \tag{3}$$

where the forget, input, and output gates regulate information flow across time steps.

The proposed model adopts a CNN–BiLSTM architecture rather than a pure LSTM, as illustrated in Figure 2. A 1D convolutional layer first extracts local temporal–spectral features from the IQ sequences. Two stacked Bidirectional LSTM layers with 64 hidden units are then applied to capture

DOI: 10.54254/2755-2721/2026.AS32102

forward and backward temporal dependencies: the first layer outputs the full sequence, while the second produces a compact temporal representation:

$$h_t = [\overrightarrow{h_t}; \overrightarrow{h_t}] \tag{4}$$

Finally, fully connected layers and a Softmax output perform modulation classification.

By integrating convolutional feature extraction with bidirectional temporal modeling, the proposed CNN–BiLSTM architecture effectively captures both local and long-range temporal dependencies in modulated signals. Specifically, the 1D convolutional layers first extract discriminative local time-domain features and suppress noise by aggregating neighboring samples, thereby providing a compact and informative representation of the input IQ sequences. The subsequent bidirectional LSTM layers model temporal dependencies in both forward and backward directions, enabling the network to capture symbol-level correlations across extended time intervals.

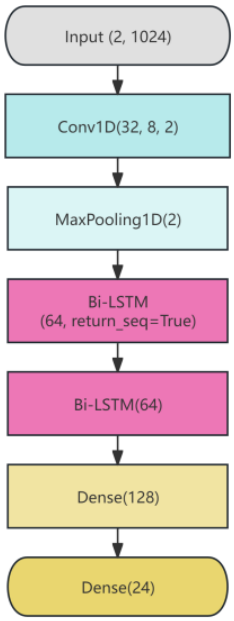

Figure 2. The architecture of CNN-BiLSTM

4. Experiment design

This study employs the RadioML 2018.01A dataset, a widely adopted benchmark for AMC [11]. The dataset consists of 24 digital and analog modulation formats, encompassing a wide range of modulation schemes commonly used in modern wireless communication systems. Representative digital modulations include BPSK, QPSK, 8PSK, 16QAM, 64QAM, and CPFSK, while analog modulations such as AM-DSB, AM-SSB, and FM are also included.

The signals are generated under diverse channel conditions with Signal-to-Noise Ratio (SNR) values ranging from −20 dB to 30 dB in 2 dB increments. For each modulation type and SNR level, an equal number of signal samples is provided, resulting in a balanced dataset across modulation categories and noise conditions. Each sample consists of complex baseband IQ data with a fixed length of 1024 samples. During data preprocessing, the complex baseband IQ samples are normalized using Z-score standardization to eliminate scale differences between channels and improve numerical stability during training. To enhance model generalization, additive white Gaussian noise (AWGN) is employed as a data augmentation strategy during training, which simulates random channel perturbations by adding zero-mean Gaussian noise to the input signals. The dataset is randomly

DOI: 10.54254/2755-2721/2026.AS32102

divided into training and testing at a 7:3 ratio. To ensure a fair comparison, both the 1D-ResNet and CNN–BiLSTM models are trained and evaluated under identical experimental conditions. Model performance is primarily evaluated using classification accuracy, with confusion matrices for detailed analysis. For the CNN–BiLSTM model, training uses the Adam optimizer with an initial learning rate of 0.0013. The batch size is set to 512, and the model is trained for a maximum of 120 epochs, with early stopping applied based on validation accuracy. For the 1D-ResNet model, the Adam optimizer is also employed, with an initial learning rate of 0.0012. The batch size is set to 128, and the maximum number of training epochs is 60, with an early stopping to prevent overfitting.

5. Result and discussion

5.1. Performance comparison

As shown in Figure 3, it compares the classification accuracy of the 1D-ResNet and CNN-BiLSTM models over the SNR range from 0 to 14 dB. At low SNR levels (0–4 dB), 1D-ResNet consistently outperforms CNN-BiLSTM. This suggests that both architectures can effectively extract discriminative features in the presence of low noise. When the SNR exceeds 10 dB, CNN-BiLSTM begins to outperform 1D-ResNet. At 10 dB, CNN-BiLSTM achieves an accuracy of 87.55%, compared to 85.45% for 1D-ResNet. These results demonstrate the superior capability of CNN-BiLSTM in exploiting long-range temporal dependencies under cleaner signal conditions.

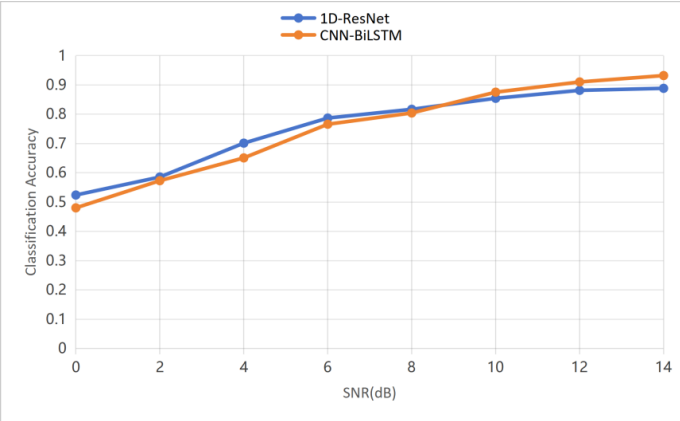

Figure 3. Classification accuracy between two architecture

5.2. Confusion matrix analysis

The normalized confusion matrices at 0 dB and 14 dB provide detailed insights into the classification behavior of 1D-ResNet and CNN-BiLSTM under different noise conditions.

At 0 dB SNR, as shown in Figure 4 and Figure 5, they show that severe noise causes notable inter-class confusion for both models, particularly among high-order modulations. However, Figure 4 indicates that 1D-ResNet exhibits stronger diagonal dominance for PSK-based modulations, achieving accuracies of 0.93 for BPSK, 0.88 for QPSK, and 0.94 for 8PSK, with minimal cross-class confusion. In contrast, as shown in Figure 5, CNN-BiLSTM suffers from increased confusion among high-order modulation schemes; for example, 64QAM achieves only 0.095 accuracy, with substantial misclassification to adjacent QAM classes. This reflects the impairment of temporal dependency modeling under strong noise.

At 14 dB SNR, as shown in Figure 6 and Figure 7, they show confusion matrices become significantly more concentrated along the diagonal. As illustrated in Figure 7, CNN-BiLSTM achieves near-perfect classification for several representative modulations, including 1.00 for OOK,

DOI: 10.54254/2755-2721/2026.AS32102

0.98 for 8PSK, and 0.95 for 16QAM, benefiting from well-preserved temporal correlations at high SNR. Meanwhile, Figure 6 shows that 1D-ResNet maintains stable recognition for structurally regular modulations, such as 1.00 for BPSK and QPSK and 0.99 for 8PSK, but exhibits mild confusion among adjacent high-order QAM classes, with 64QAM achieving about 0.38 accuracy.

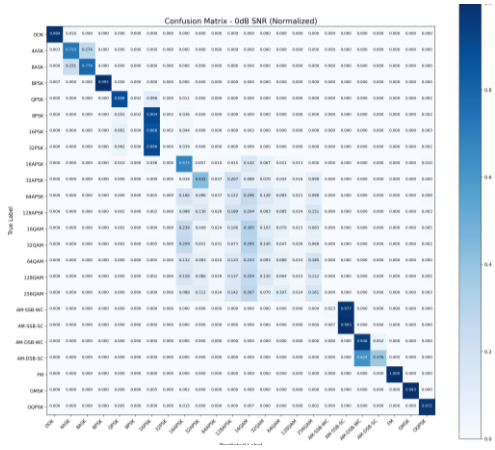

Figure 4. 1D-ResNet 0dB

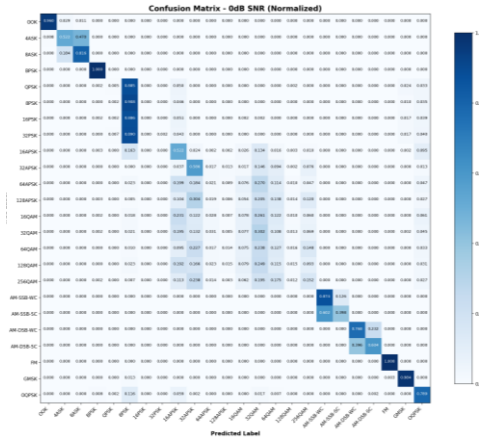

Figure 5. CNN-BiLSTM 0dB

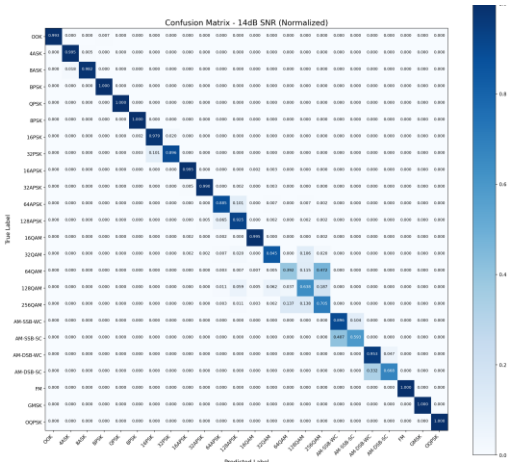

Figure 6. 1D-ResNet 14dB

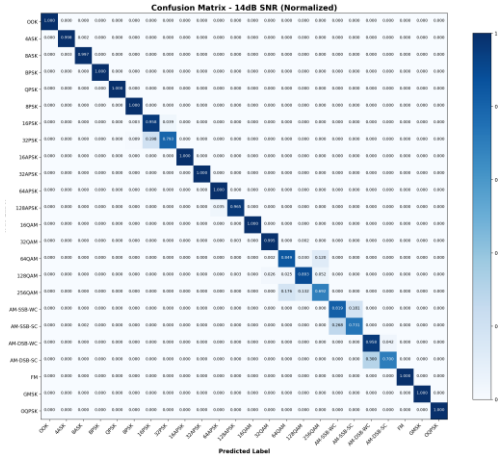

Figure 7. CNN-BiLSTM 14dB

5.3. Discussion on model complementarity

Experimental results reveal a strong complementarity between 1D-ResNet and CNN-BiLSTM. 1D-ResNet excels at local feature extraction from I/Q samples, making it more effective under low-SNR conditions. CNN-BiLSTM leverages global temporal modeling to achieve superior performance at medium-to-high SNR levels. This complementary behavior indicates that a hybrid architecture combining 1D-ResNet and CNN-BiLSTM could further enhance AMC performance across diverse channel conditions. A sequential fusion strategy can be adopted: 1D-ResNet first performs robust local feature extraction from raw I/Q samples, particularly under low-SNR conditions. High-level representations extracted by the residual network are then fed into a CNN-BiLSTM module to model long-range temporal dependencies. Integrating noise-resilient spatial features with temporal context modeling, this cascaded architecture is expected to achieve improved classification accuracy across a wide SNR range.

DOI: 10.54254/2755-2721/2026.AS32102

6. Conclusion

The experimental results demonstrate that the two models exhibit distinct and complementary advantages under different noise environment. Specifically, 1D-ResNet consistently outperforms CNN–BiLSTM in low-SNR scenarios (0-4 dB), benefiting from robust hierarchical structural feature extraction and strong resistance to noise degradation. In contrast, CNN–BiLSTM achieves superior performance in mid-to-high SNR conditions (≥10 dB) by effectively modeling long-term temporal dependencies in the signal sequence, with 10 dB identified as a critical performance transition point between the two architectures. Furthermore, CNN–BiLSTM shows particular advantages in low- and medium-order modulation schemes, whereas 1D-ResNet demonstrates stronger discrimination capability for high-order PSK modulations. Despite these promising results, several limitations remain. First, the evaluated SNR range is limited to 0–14 dB, excluding extreme operating conditions such as ultra-low (<0 dB) or very high (>14 dB) SNR regimes. Second, this study focuses solely on conventional digital modulation formats, without considering more complex or emerging waveform designs. Third, the experiments rely on simulated datasets, which cannot fully capture real-world impairments like severe multipath fading, hardware nonlinearities, and external interference. First, the evaluated SNR range is restricted to 0–14 dB and does not cover extreme operating conditions such as ultra-low (<0 dB) or very high (>14 dB) SNR regimes. Second, this study focuses exclusively on conventional digital modulation formats, without considering more complex or emerging waveform designs. Future work will address these limitations by extending the evaluation to broader and extreme SNR conditions, incorporating more diverse modulation schemes, and validating the proposed findings using real-world over-the-air datasets.

References

[1] Jdid, B., Hassan, K., Dayoub, I., Lim, W. H., & Mokayef, M. (2021). Machine learning based automatic modulation recognition for wireless communications: A comprehensive survey. IEEE Access, 9, 57851-57873.

[2] Yang, C., He, Z., Peng, Y., Wang, Y., & Yang, J. (2019). Deep learning aided method for automatic modulation recognition. IEEE Access, 7, 109063-109068.

[3] Ansari, S., Mahmoud, S., Majzoub, S., Almajali, E., Jarndal, A., & Bonny, T. (2025). A Novel LSTM Architecture for Automatic Modulation Recognition: Comparative Analysis with Conventional Machine Learning and RNN-Based Approaches. IEEE Access.

[4] Le, H.-K., Hoang, V.-P., & Doan, V.-S. (2024). AMC-CNN-BiLSTM: An advanced hybrid model for automatic modulation classification of radio signals. In 2024 RIVF International Conference on Computing and Communication Technologies (RIVF) (pp. 271–275). IEEE.

[5] Hamza, M. A., Hassine, S. B., Larabi-Marie-Sainte, S., Nour, M. K., Al-Wesabi, F. N., Motwakel, A., ... & Al Duhayyim, M. (2022). Optimal bidirectional LSTM for modulation signal classification in communication systems. Cmc-Comput. Mater. Contin, 72, 3055-3071.

[6] Vijayamohanan, J., Gupta, A., & Christodoulou, C. (2025). Joint source detection and modulation classification with neural networks. In 2025 IEEE International Symposium on Antennas and Propagation and North American Radio Science Meeting (AP-S/CNC-USNC-URSI) (pp. 1803–1805). IEEE.

[7] Lu, X., Tao, M., et al. (2021). Lightweight network design based on ResNet structure for modulation recognition. In 2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall) (pp. 1-5). IEEE.

[8] Choi, S., Chung, S., Lee, S., Han, S., Kang, T., Seo, J., Kwak, I.-Y., & Oh, S. (2024). TB-ResNet: Bridging the gap from TDNN to ResNet in automatic speaker verification with temporal-bottleneck enhancement. In 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 10291–10295). IEEE.

[9] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[10] S. Hochreiter and J. Schmidhuber, 1997. "Long Short-Term Memory," in Neural Computation, vol. 9, no. 8, pp. 1735-1780, 15 Nov. doi: 10.1162/neco.1997.9.8.1735.

[11] O'Shea, T. J., Roy, T., & Clancy, T. C. (2018). Over-the-air deep learning based radio signal classification. IEEE Journal of Selected Topics in Signal Processing, 12(1), 168-179.