

Research of Automatic Modulation Recognition Technology Based on Deep Learning

Nanbin Liu

*School of Electronic information technology, Beijing Jiaotong University, Beijing, China
23723054@bjtu.edu.cn*

Abstract: Automatic modulation recognition is a vital component in communication systems. It helps in recognizing and classifying signal' modulation automatically without the acquirement of the signal's modulation type in advance. This paper aims to provide a comprehensive exploration of the application of deep learning techniques in the field of Automatic Modulation Recognition (AMR), with a specific emphasis on the analysis and comparison of two prominent neural network architectures: Convolutional Neural Networks (CNN) and Convolutional Long Short-Term Memory Networks (CLDNN). By delving into the structural and functional aspects of these models, the paper seeks to elucidate how CNNs, with their ability to capture spatial features through convolutional layers, and CLDNNs, which enhance this capability by integrating Long Short-Term Memory (LSTM) units to handle temporal dynamics, contribute to the advancement of AMR technology. We used the DeepSig dataset: RadioML 2018.01A to evaluate them, comparing their performance in different signal-to-noise ratio scenarios. The results manifested that CLDNN performed better than CNN obviously, especially in the low SNR scenario where CLDNN reached a higher validation set accuracy.

Keywords: Deep learning, CNN, CLDNN, AMR, communication engineering

1. Introduction

Nowadays, transmitted data is becoming more complex and heavier, mainly due to the rapid development of wireless communication technology. Automatic modulation recognition (AMR) excels in addressing the challenge of identifying the modulation type of incoming signals without prior knowledge of the transmitted information, thereby empowering receiver systems with enhanced capabilities. AMR is indispensable in spectrum management, cognitive radio, and military applications, as its implementation significantly boosts the security and efficiency of information transmission. In the military domain, AMR technology enables the effective identification of key parameters and features of adversary communication signals, facilitating the anticipation of enemy decisions and providing a strategic advantage in electronic countermeasures. Within the civilian sector, AMR aids radio management authorities in the efficient regulation of radio resources, curbing frequency occupation by unauthorized broadcasters, and ensuring the integrity and clarity of standard communication channels. However, traditional methods such as feature-s-based and likelihood-based approaches show difficulties in low-SNR conditions. The approach based on maximum likelihood has high computational complexity and depends on the specific parameters of the transmitter. Additionally, these methods often depend on manual feature extraction, which not only requires

professional knowledge but also risks overlooking critical features, such as those in the joint time and frequency domains. With the advancement of Deep Learning (DL), models like Convolutional Neural Networks (CNN) are demonstrating great potential in surmounting these challenges. DL models exhibit superior adaptability and performance in complex, low-SNR environments. Zhang's group in 2023 [1] announced that deep learning-based models are able to study robust feature representations, improving the classification accuracy without prior professional knowledge. Modulation recognition is becoming a current mission in modern communication systems with the increasing need for spectrum. AMR's importance extends to civil applications, such as cognitive radio, spectrum sensing, and software-defined radio, and military operations, including electronic warfare and secure communications [2]. Despite its significance, traditional AMR methods typically perform poorly in low-SNR conditions, where signal interference can lead to unclear reception, complicating accurate signal classification. Consequently, integrating DL into AMR technology to bolster its robustness in low-SNR scenarios represents a critical research focus. This paper primarily investigates the effectiveness of DL in AMR. For this purpose, the DeepSig Dataset: 2018.01A, which includes various types of labeled signals, is utilized for training and testing models. The study employs models based on CNN and CLDNN architectures, and experiments are conducted on a personal computer across four different SNR conditions (0 dB, 4 dB, 10 dB, 16 dB) to evaluate the classification accuracy of these models under low SNR levels.

2. Methodology

This section will introduce the detailed procedures in this research, including the dataset, preprocessing steps, models structure and training procedures. This research will focus on 2 models: CNN and CLDNN. After that, this study will give the test results and evaluation metrics.

2.1. Dataset and preprocessing

Dataset used in the research is DeepSig Dataset: RadioML 2018.01A, which contains I/O signals modulated in 24 modulations: OOK, ASK4, ASK8, BPSK, QPSK, PSK8, PSK16, PSK32, APSK16, APSK32, APSK64, APSK128, QAM16, QAM32, QAM64, QAM128, QAM256, AM_SSB_WC, AM_SSB_SC, AM_DSB_WC, AM_DSB_SC, FM, GMSK and OQPS. To further enhance the robustness and accuracy of the models during training, the dataset undergoes preprocessing, which includes normalization, segmentation, and augmentation. Preprocessing is crucial for ensuring the models' generalization capability in dynamic environments characterized by high levels of variable noise.

2.2. Convolutional neural network (CNN) architecture

The CNN model efficiently extracts specific features from IQ data. The CNN architecture comprises multiple convolutional layers, each followed by batch normalization, ReLU activation, and max-pooling operations. These components operate in sequence, ensuring that the network captures essential modulation characteristics while reducing computational complexity. Here are the main functions of the layers offered.

Convolutional layers: These layers capture the modulation scheme of local patterns through using the 2D convolutional filter to process the IQ data.

Pooling layers: These layers reduce spatial dimensions while preserving fundamental features and minimizing the risk of overfitting.

Dense layer: Following the extraction of characteristics, the dense layer performs the final classification, which is then followed by a softmax function to produce probabilistic outputs. The

hybrid model is designed to capture both spatial and temporal features, thereby enhancing robustness in low-SNR environments.

2.3. Convolutional Long-Short Term Neural Network (CLDNN) architecture

Convolutional Long-Short Term Neural Networks (CLDNNs) integrate CNN and LSTM models, inheriting the CNN's capability to extract features and the LSTM's power for temporal modeling. The main layers and their functions are as follows:

CNN Layer: Extracts spatial features from input data through convolutional operations.

LSTM Layer: Captures temporal characteristics from the data output by the CNN layer. This layer is particularly efficient for signals with sequential modulation patterns and dynamic channel variations [3].

Dense Layer (Fully Connected Layer): Processes the features extracted by the CNN and LSTM layers to classify and output probabilities through Softmax function activation.

Configuration: CNN and CLDNN models both use Adam optimizer.

Operating: These 2 models were set to run for 100 epochs in every 4 kinds of SNR conditions (0, 4, 10, 16dB) with patience is set to 50 for CNN and 100 for CLDNN. Nevertheless, CLDNN and CNN are still not able to normally run to 100 epochs.

3. Results and discussion

This sector will evaluate the performance of CNN and CLDNN in AMR according to several metrics, including snr-accuracy curves and confusion metrics.

3.1. Performance under each SNR level

Models are trained using the sliced DeepSig Dataset: RadioML 2018.01A, under SNR conditions of 0 dB, 4 dB, 10 dB, and 16 dB. The accuracy of each model is tested on the validation set for each SNR environment. Accuracy is evaluated using confusion matrices, as follows:

$$\text{Accuracy} = \frac{\sum \text{Diagonal elements}}{\sum \text{Total matrix elements}} \quad (1)$$

The records of accuracy are summarized into a line graph in Figure 1.

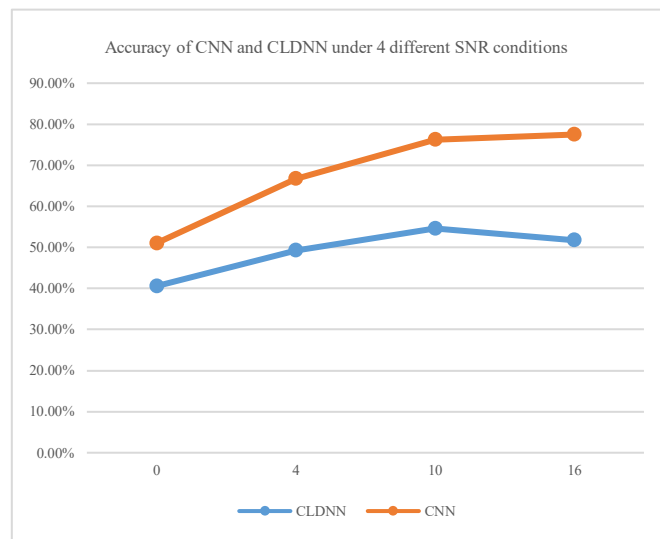


Figure 1: Accuracy of CNN and CLDNN under 0, 4,10,16dB snr environments

3.2. Confusion matrix

Confusion matrixes are generated after each model is trained 100 epochs across 4 different SNR environments. Elements in the matrix represent the probability that a modulation type on the vertical axis will be recognized as a modulation type on the horizontal axis. The confusion matrices presented in Figures 2-9 indicate that the CNN excels in distinguishing high-order QAM modulation schemes, whereas the CLDNN demonstrates exceptional robustness against PSK modulation interference. These confusion matrices are generated after each model has been trained for 100 epochs across four different SNR environments. They serve as a valuable reference for evaluating the performance of the models.

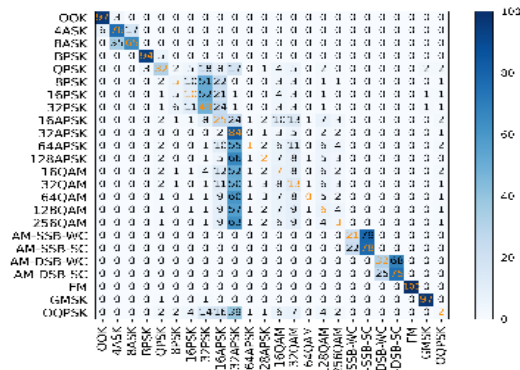


Figure 2: Confusion matrix of CLDNN in 0dB

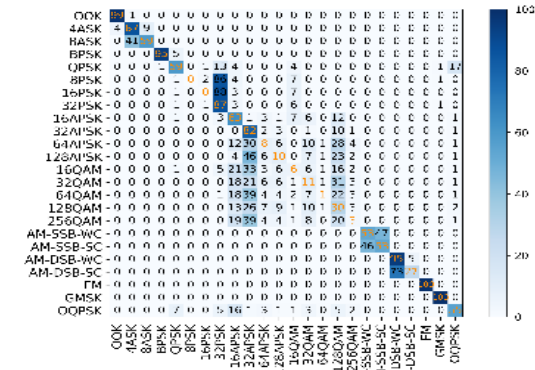


Figure 3: Confusion matrix of CLDNN in 4dB

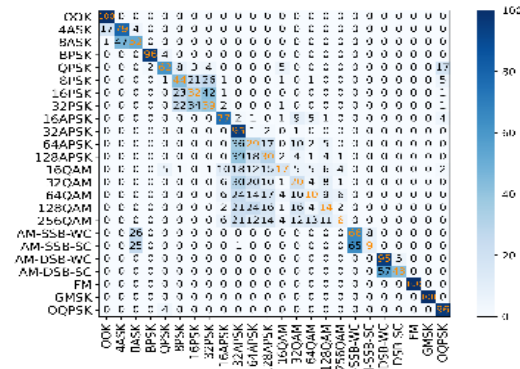


Figure 4: Confusion matrix of CLDNN in 10dB

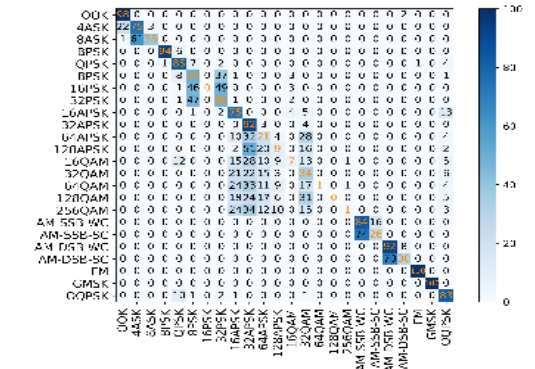


Figure 5: Confusion matrix of CLDNN in 16dB

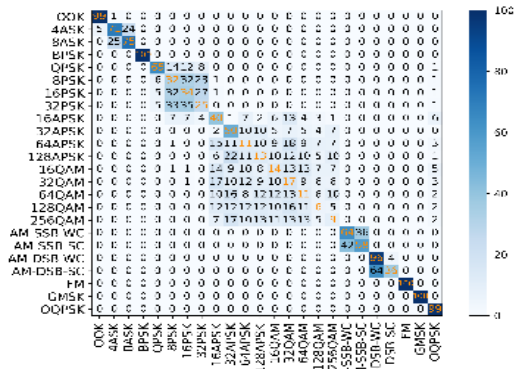


Figure 6: Confusion matrix of CNN in 0dB

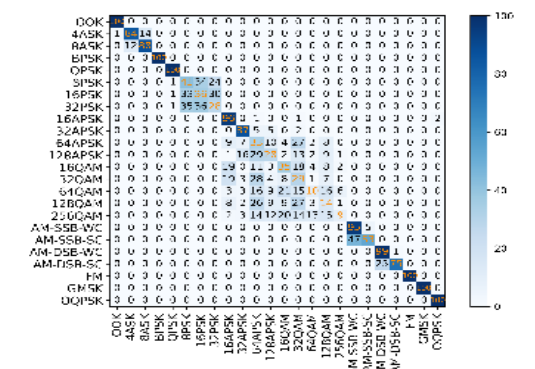


Figure 7: Confusion matrix of CNN in 4dB

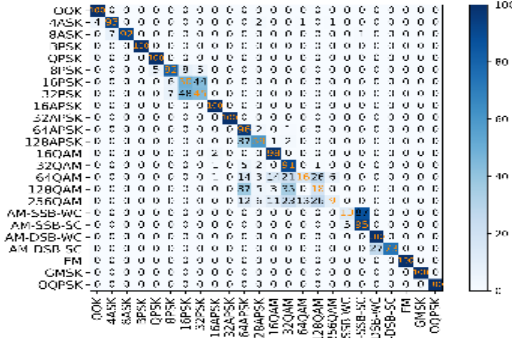


Figure 8: Confusion matrix of CNN in 10dB

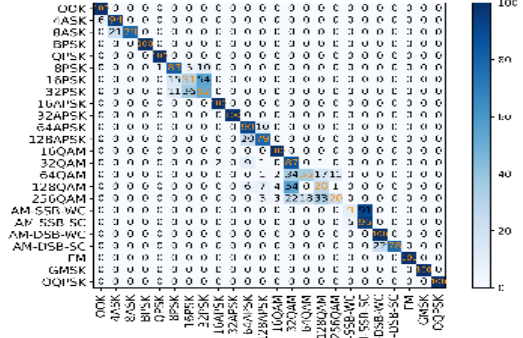


Figure 9: Confusion matrix of CNN in 16dB

The confusion matrices presented in Figures 2-9 indicate that the CNN demonstrated superior performance in distinguishing high-order QAM modulation schemes, while the CLDNN exhibited excellent robustness against PSK modulation interference. Additionally, it was observed that the CNN achieved better performance in the recognition of most modulation types, as evidenced by a higher accuracy in the confusion matrices compared to the CLDNN.

There is another thing need to be noticed, in the process of distinguishing AM-SSB-WC and AM-SSB-SC in CNN, there is still a problem that AM-SSB-WC is wrongly identified as AM-SSB-SC with a high probability under the snr of 16dB, but most of the AM-SSB-SC are successfully identified. However, in the process of distinguishing AM-SSB-WC and AM-SSB-SC in CLDNN, AM-SSB-SC has a high probability of being wrongly identified as AM-SSB-WC, but most AM-SSB-WC are successfully identified. At the same time, in the process of CLDNN distinguishing AM-DSB-SC and AM-DSB-WC, there is also a problem that AM-DSB-SC has a high probability of being wrongly identified as AM-DSB-WC at 16dB SNR, but most AM-DSB-WC are successfully identified.

3.3. Results analysis

This section will discuss some of the underlying reasons behind the observed results.

We have observed that CNN performed better than CLDNN in some cases. CNN is not as sensitive as CLDNN to temporal dependence. Consequently, the CNN exhibited greater robustness across varying SNR conditions. In contrast, the CLDNN, due to its reliance on sequential signal information, was more susceptible to interference, particularly in low SNR environments. This susceptibility necessitated the intervention to terminate the CLDNN training process before reaching 100 epochs.

These results reveal persistent classification ambiguities between AM-SSB-WC and AM-SSB-SC under 16dB SNR conditions. For CNN architectures, misclassification occurs where AM-SSB-WC signals are incorrectly identified as AM-SSB-SC counterparts, despite sufficient signal clarity (10/16dB SNR). This phenomenon likely stems from spectral signature overlaps induced by the residual carrier component in AM-SSB-WC waveforms, which diminishes the inter-class discriminability in frequency-domain representations. Notably, CNN demonstrates effective capture of AM-SSB-SC's distinctive spectral null characteristics at higher SNR levels, achieving reliable identification when carrier suppression artifacts become pronounced. CLDNN exhibits an inverse confusion pattern, preferentially misclassifying AM-SSB-SC as AM-SSB-WC while maintaining high precision for AM-SSB-WC detection. This inversion suggests CLDNN's heavy reliance on temporal feature extraction through its LSTM modules - the suppressed carrier in AM-SSB-SC introduces transient amplitude fluctuations that may be misinterpreted as noise-induced temporal variations, whereas AM-SSB-WC's stable carrier component provides clearer time-domain reference points. Similar cross-classification challenges emerge in AM-DSB-SC/WC differentiation, where

CLDNN struggles with spectral similarity despite 16dB SNR conditions. The dual sideband structures and symmetrical spectral components in both DSB variants create ambiguous time-frequency representations that confuse temporal modeling pathways. These observations underscore fundamental limitations in current DL architectures: CNNs' spectral bias proves vulnerable to carrier-induced feature overlaps, while CLDNNs' temporal sensitivity amplifies interpretation errors in signals with suppressed carriers. The persistent confusion at 16dB SNR indicates that neither spatial nor temporal feature extraction alone suffices for discriminating carrier-modulated twin signals. Potential mitigation strategies could involve hybrid architectures combining parallel spectral-temporal analysis streams, or adversarial training with carrier suppression artifacts augmentation to strengthen model discernment. Implementing attention mechanisms to dynamically weight critical spectral regions might further enhance differentiation capacity for these high-similarity modulation pairs.

Furthermore, experimental observations indicate diminished classification efficacy of CLDNN architectures at 10dB SNR compared to 16dB SNR conditions. This degradation principally stems from the model's inherent dependency on temporal coherence analysis—a mechanism increasingly vulnerable to stochastic noise perturbations at moderate SNR levels. The LSTM-driven temporal feature extraction becomes suboptimal when ambient noise obscures subtle amplitude/phase transitions critical for sequence modeling. At elevated SNR (16dB), enhanced signal integrity allows CLDNN to resolve carrier synchronization patterns and transient envelope variations with greater fidelity, thereby recovering its temporal modeling advantages. The performance disparity underscores a critical trade-off: while CLDNN excels in high-SNR temporal dynamics interpretation, its noise susceptibility at intermediate SNR ranges necessitates architectural refinements for robust operation across variable channel conditions. Implementing noise-robust temporal attention mechanisms or hybrid spectral-temporal fusion layers could potentially mitigate this limitation.

3.4. Future works

3.4.1. Enhanced temporal modeling based on attention mechanism

The difficulty demonstrated in CLDNN manifests the benefits of attention mechanism. By applying attention layers, the model could focus more on the most relevant temporal features, improving its ability to differentiate between signals with similar spectral and temporal characteristics, and this approach really works in the high snr conditions, where the temporal structure plays a critical role in modulation recognition. Wang and Zhao (2021) [4] demonstrated that attention-based neural networks are effective in focusing on the most important temporal features of the signal, which could help mitigate issues of misclassification, especially in high-noise environments where subtle differences in modulation types are harder to detect. Furthermore, recent studies, such as LeCun et al. (2015) [5], have shown that incorporating deep learning techniques, including attention mechanisms, can substantially enhance the ability to capture critical temporal features in complex datasets.

3.4.2. Data Augmentation and Class Balancing

Another approach to improving model performance is utilizing data augmentation techniques and balancing the training dataset. Future research should explore data augmentation techniques, as suggested by Qian and Sun (2023) [6], where synthetic data generation or adversarial training can be used to balance the dataset and provide the model with a more comprehensive understanding of less frequent modulation schemes. Augmenting the dataset through methods like synthetic data generation or adversarial training could help the model learn a broader range of features and reduce misclassifications, especially in the cases that some modulation schemes are not fully underrepresented.

3.4.3. Improving Feature Extraction with Frequency-Domain Techniques

Considering the spectrum overlapping between modulation schemes such as AM-SSB-WC and AM-SSB-SC using techniques such as Short-Time Fourier Transform (STFT) or wavelet transforms might offer richer feature sets, incorporating additional frequency-domain features or enhancing the frequency analysis part of the model.

3.4.4. Adaptive Training Strategies

There exist some cases that the premature stopping observed in CLDNN, adaptive training strategies can be implemented. Dynamic adjustment of patience or learning rate could allow the model to adaptively converge to an optimal solution without prematurely halting the learning process. This strategy would help both CNN and CLDNN achieve better generalization, particularly when the models are exposed to noisy or complex signal patterns.

4. Conclusion

This research investigates the performance of two deep learning models—CNN and CLDNN—in automatic modulation recognition (AMR) across a series of SNR environments. Although the CNN demonstrates excellent performance in high SNR conditions, it still exhibits confusion with certain closely related modulation schemes, such as AM-SSB-WC and AM-SSB-SC. This confusion can be attributed to the spectral similarities between these two modulation types, where the presence of the carrier in AM-SSB-WC overlaps with the spectral features of AM-SSB-SC. This suggests that the model is more adept at recognizing certain modulation types when spectral distinctions are more pronounced. In contrast, CLDNN combines the convolution layers with temporal modeling of LSTM. Nevertheless, in classifying AM-SSB-WC and AM-SSB-SC, AM-SSB-WC is largely classified correctly, while AM-SSB-SC is frequently misidentified as AM-SSB-WC. This indicates the model's sensitivity to temporal features, where AM-SSB-SC's temporal structure may be overshadowed by the stable time-series characteristics of AM-SSB-WC. Additionally, CLDNN encounters similar issues when distinguishing between AM-DSB-SC and AM-DSB-WC at 16dB SNR, with AM-DSB-SC often being misclassified as AM-DSB-WC. The overlapping spectral features and the temporal modeling dependencies contribute to these misclassifications. These results underline the difficulties in distinguishing the modulation schemes that have similar spectral or temporal characteristics even in high SNR conditions. Despite these challenges, both models offer valuable insights: CNN excels at capturing spatial features, particularly under higher SNR conditions, while CLDNN provides an advantage by modeling temporal dependencies, making it effective in environments where sequential information is crucial. However, both models need to be further optimized to improve classification accuracy in more complex scenarios.

References

- [1] Zhang, F. (2023). Research on automatic modulation recognition technology based on deep learning. *Journal of Communication Technology and Applications*, 18(4), 123-131.
- [2] Xu, C., Zhang, Y., & Li, Q. (2023). Research on modulation recognition based on deep learning under low signal-to-noise ratio. *Wireless Communications Journal*, 12(3), 56-64.
- [3] Emam, A., Shalaby, M., Aboelazm, M. A., Bakr, H. E. A. & Mansour H. A. A. (2020) A Comparative Study between CNN, LSTM, and CLDNN Models in The Context of Radio Modulation Classification, 2020 12th International Conference on Electrical Engineering (ICEENG), Cairo, Egypt, 2020, pp. 190-195, doi: 10.1109/ICEENG45378.2020.9171706.
- [4] Wang, Z., & Zhao, Y. (2021). Attention-based deep learning for signal classification. *Journal of Communication Technology*, 45(3), 189-200.
- [5] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.

- [6] Qian, H., & Sun, X. (2023). *Exploring data augmentation methods for class imbalance in signal processing*. *IEEE Transactions on Signal Processing*, 71, 129–141.