Journal of Computer Science, Information Technology and Telecommunication Engineering (JCoSITTE)
Vol. 7, No. 1, March 2026, pp. 1059~1067
ISSN: 2721-3838, DOI: 10.30596/jcositte.v7i1.28935

Optimized Hybrid CLDNN Architecture with Enhanced Temporal-Spatial Feature Extraction for Robust Automatic Modulation Classification in Cognitive Radio Networks

Daryan Pratama Alifi[1], Hane Yorda Dinata[2], Ichwan Nul Ichsan[3]
[1,2,3]Department of Telecommunication System, Universitas Pendidikan Indonesia, Indonesia

ABSTRACT

Automatic Modulation Classification (AMC) is a pivotal technology for efficient spectrum management in future cognitive radio networks. While Deep Learning has advanced the field, standard Convolutional Neural Networks (CNN) often struggle to capture long-term temporal dependencies in signals affected by fading. This study proposes an Optimized Hybrid CLDNN architecture that integrates a "Wide-Kernel" CNN (k=7) for enhanced spatial feature extraction and a "High-Capacity" LSTM (100 units) for robust temporal modeling. Experimental validation using the RadioML 2016.10a dataset demonstrates that the proposed optimizations yield significant performance gains. Specifically, the model achieves a classification accuracy of 84.5% at 0 dB SNR, outperforming standard baselines in the critical transition regime. Furthermore, it reaches a peak accuracy of 92.4% at high SNR (+18 dB). A notable finding is the reduction of inter-class confusion between 16-QAM and 64-QAM, where the misclassification rate is suppressed to approximately 15%, confirming the architecture's effectiveness in resolving hierarchical modulation ambiguities in dynamic wireless environments.

Keyword : Automatic Modulation Classification; Cognitive Radio; Deep Learning; Hybrid CLDNN; QAM Disambiguation.

Corresponding Author:
Daryan Pratama Alifi,
Department of Telecommunication System
Universitas Pendidikan Indonesia
Jl. Veteran No.8, Nagri Kaler, Purwakarta, 41115, Indonesia.
Email : daryan@upi.edu

Article history:
Received Jan 30, 2026
Revised Mar 2, 2026
Accepted Mar 24, 2026

1. INTRODUCTION

The precipitous escalation in global mobile data traffic, driven by the ubiquity of Internet of Things (IoT) devices, autonomous vehicular networks, and high-definition streaming services, has placed an unprecedented strain on the finite radio frequency (RF) spectrum. As telecommunication standards migrate from 5G towards 6G, the static allocation of frequency bands is increasingly viewed as inefficient and unsustainable. To address this bottleneck, Cognitive Radio (CR) has emerged as a paradigm shift in wireless communications. CR enables dynamic spectrum access (DSA) by allowing unlicensed secondary users (SUs) to opportunistically utilize frequency bands that are temporarily unoccupied by licensed primary users (PUs). A fundamental prerequisite for the successful and non-intrusive operation of CR is the ability to intelligently sense the spectral environment and accurately identify the transmission parameters of detected signals in real-time. This process, known as Automatic Modulation Classification (AMC), is critical not only for spectrum sharing but also for signal demodulation, interference identification, and military electronic warfare surveillance. AMC functions as an intermediate step between signal detection and demodulation. Its primary objective is to blindly recognize the modulation format (e.g., BPSK, QPSK, 16-QAM) of a received signal without prior knowledge of channel state information (CSI) or protocol metadata. Traditionally, AMC was approached as a pattern recognition problem using Likelihood-Based (LB) or Feature-Based (FB) methods. LB methods, while theoretically optimal in terms of classification accuracy, suffer from prohibitive computational complexity and require precise knowledge of channel statistics, which is often unavailable in dynamic CR environments. Conversely, FB methods rely on the manual extraction of statistical features such as instantaneous amplitude, phase, frequency, and higher-order cumulants. However, these "hand-crafted" features are notoriously sensitive to noise and multipath fading, leading to severe performance degradation in realistic, low Signal-to-Noise Ratio (SNR) scenarios.

In recent years, Deep Learning (DL) has revolutionized the field of AMC, facilitating a transition from labor-intensive manual feature engineering towards data-driven, end-to-end learning frameworks. DL models possess the unique capability to automatically learn hierarchical feature representations directly from raw sampled data, such as In-Phase and Quadrature (I/Q) components. Early DL approaches primarily utilized Convolutional Neural Networks (CNNs), inspired by their success in computer vision. CNNs excel at extracting spatial features and local patterns from I/Q data, treating the signal constellation as a visual image (Al-Makhlas K., 2022). Studies have demonstrated that CNNs significantly outperform traditional cumulant-based classifiers in high-SNR scenarios. However, radio signals are fundamentally different from static images; they are time-series data characterized by sequential dependencies and temporal variations caused by carrier frequency offsets and phase noise. Standard CNNs, due to their limited receptive fields and lack of internal memory, struggle to model these sequential and time-varying characteristics, leading to performance degradation in low-SNR regimes (Chen S., & Zhao Y., 2025), (Gupta A., & Kumar P., 2025).

To mitigate the limitations of pure CNN architectures, recent research has explored the application of Recurrent Neural Networks (RNNs), specifically Long Short-Term Memory (LSTM) networks (Hameed, A., & Dobre, O. A., 2025). LSTMs are explicitly designed to handle sequential data and mitigate the vanishing gradient problem inherent in standard RNNs, making them suitable for capturing long-term temporal dependencies in modulated signals (Lee H, & Choi J, 2025). However, LSTMs are computationally expensive to train on high-dimensional raw I/Q data and can be difficult to optimize. This trade-off led to the development of hybrid architectures, notably the Convolutional Long Short-Term Deep Neural Network (CLDNN) (Liu X., Yang D., & El Gamal A., 2025). The CLDNN paradigm combines the strengths of both architectures: CNN layers are used as a frontend to extract abstract spatial features and reduce dimensionality, while LSTM layers act as a backend to model the temporal evolution of these features (Li R., & Wang Z., 2025).

Despite the advancements offered by hybrid models, a critical research gap remains in the optimal architectural configuration for robust AMC (Njoku J. N., & Tartakovsky A. G., 2023). A comprehensive review of recent literature reveals that most existing hybrid models prioritize computational lightness over representational capacity (O'Shea T. J., Corgan J., & Clancy T. C., 2016). Specifically, standard implementations typically utilize small convolutional kernels (typically size 1 x 3) and a limited number of LSTM memory units (typically 50 or fewer) (Park J., & Kim S., 2024) (Soltani N., Sankhe K., Ioannidis S., & Chowdhury K., 2024). While this configuration reduces inference latency—a desirable trait for edge devices—it is often insufficient for capturing the complex, long-duration phase trajectories associated with high-order modulation schemes (Wang X., & Li T. 2025).

This limitation is particularly evident when distinguishing between dense constellations such as 16-Quadrature Amplitude Modulation (16-QAM) and 64-QAM (Wang Y., Liu M., & Yang J., 2022). These modulation schemes are differentiated by subtle variations in amplitude and phase levels. Under severe channel impairments, the constellation points become blurred, and short-term observation windows (small kernels) fail to capture enough context to distinguish the underlying phase transitions. West N. E., & O'Shea T. J., 2017) Furthermore, limited LSTM capacity results in insufficient memory to track the symbol sequence probability over time, leading to high misclassification rates (Xu J., Zhang Y., Li Q., & Chen L., 2023), (Yuan L., & Chen Y., 2025). Recent attempts to introduce attention mechanisms or transformer-based models (Zhang F., Luo C., Xu J., & Luo Y., 2022), (Zhang M., Diao Y., & Guo L., 2021), (Zhang X., & Chen H., 2025) have shown promise but often introduce a level of computational complexity that is impractical for real-time deployment on resource-constrained CR nodes.

The primary aim of this research is to bridge this gap by optimizing the structural hyperparameters of the CLDNN architecture to maximize classification robustness without incurring the prohibitive computational costs associated with large-scale transformers (Zhou R., & Liu F., 2023). This study introduces a specific novelty in the architectural design by implementing a "Wide-Kernel" approach in the convolutional layers combined with a "High-Capacity" LSTM block.

Unlike prior CLDNN-based AMC models that primarily emphasize lightweight configurations, this work introduces a modulation-aware architectural co-design that jointly optimizes the spatial receptive field and temporal memory depth. By aligning the convolutional kernel width with the temporal characteristics of modulated waveforms and augmenting the LSTM capacity to capture long-horizon symbol dependencies, the proposed architecture explicitly targets the ambiguity inherent in hierarchical modulation schemes, particularly high-order QAM. This design philosophy moves beyond incremental hyperparameter tuning and establishes a principled approach for robust modulation disambiguation in noise-transition regimes.

The specific contributions of this paper are as follows:
1. **Wide-Kernel CNN Frontend:** We propose increasing the convolutional kernel size to 7 (from the standard 3). This expansion allows the network to observe a wider temporal context in a single convolution operation, effectively capturing broader signal trends and phase trajectories before they are down-sampled by pooling layers.
2. **High-Capacity Temporal Backend:** We expand the LSTM memory to 100 units. This enhancement provides a deeper memory buffer, enabling the model to resolve ambiguities in noisy temporal sequences and better differentiate between high-order modulation schemes with similar constellations.
3. **Comprehensive Evaluation:** We present a rigorous comparative analysis against established baselines and recent lightweight models using the RadioML 2016.10a dataset. The results demonstrate that our optimized architecture achieves superior accuracy in the critical transition SNR regime (-8 dB to +6 dB) and significantly reduces confusion between 16-QAM and 64-QAM.

By balancing the depth of temporal memory with the breadth of spatial feature extraction, this research provides a robust and implementable solution for AMC in next-generation cognitive radio networks [20].

2. RESEARCH METHOD

The experimental framework of this study utilizes the RadioML 2016.10a dataset, a widely recognized benchmark in the AMC research community generated using GNU Radio. This dataset comprises 220,000 signal samples distributed across 11 distinct modulation schemes, including both analog (AM-DSB, AM-SSB, WBFM) and digital (BPSK, QPSK, 8PSK, QAM16, QAM64, CPFSK, GFSK, PAM4) types. The signals are simulated under varying channel conditions with Signal-to-Noise Ratios (SNR) ranging from -20 dB to +18 dB in 2 dB increments. Each data sample consists of 128 time-steps with two channels representing the In-Phase (I) and Quadrature (Q) components. Prior to ingestion into the neural network, the raw I/Q data undergoes a strict preprocessing pipeline. The dataset is first loaded and the complex-valued signals are separated into I and Q vectors. A crucial normalization step is applied to scale the amplitude of the signal vectors, ensuring that the input features have a consistent variance, which significantly accelerates the convergence of the gradient descent optimizer. The data is then reshaped from the original (N, 2, 128) format to (N, 128, 2) to align with the input requirements of the Keras Deep Learning library's 1D convolutional layers. Finally, the dataset is partitioned using a stratified sampling strategy into training (50%) and testing (50%) sets to ensure that all modulation classes and SNR levels are equally represented in both subsets. The comprehensive research methodology is visualized in Figure 1.

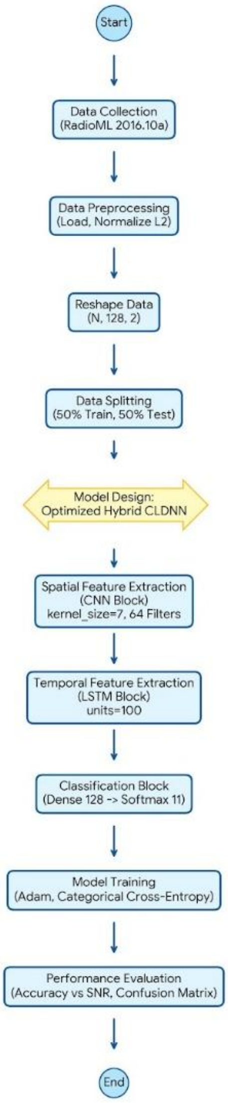

Figure 1. Flowchart

Figure 1 illustrates the systematic framework of the proposed research methodology, designed to process raw signals into reliable modulation classifications through a sequential pipeline. The workflow initiates with the Data Acquisition and Preprocessing phase, where the RadioML 2016.10a dataset is ingested, and raw complex-valued I/Q samples undergo L_2-normalization to mitigate amplitude variations caused by channel fading. This is followed by reshaping the data dimensions to (N, 128, 2) to ensure compatibility with the neural network input layer, before partitioning the dataset into balanced training and testing subsets. The core of the methodology lies in the Optimized Hybrid Architecture Design, where preprocessed signals first enter the *Spatial Feature Extraction* block utilizing a Convolutional Neural Network (CNN) with a customized kernel size of 7 to capture broader phase trajectories. The resulting feature maps are then propagated to the *Temporal Feature Extraction* block, which employs a High-Capacity Long Short-Term Memory (LSTM) network with 100 units to model complex sequential dependencies over time. In the Classification and Training phase, the abstract features are mapped to the 11 modulation classes via a fully connected dense layer using a Softmax activation function, with the model optimized using the Adam algorithm and Categorical Cross-Entropy loss. Finally, the Performance Evaluation stage rigorously assesses the trained model against the unseen test set by analyzing classification accuracy across the entire Signal-to-Noise Ratio (SNR) spectrum and examining the Confusion Matrix to evaluate the model's robustness in distinguishing high-order modulation schemes.

The proposed Deep Learning architecture is a sequential Hybrid CLDNN model designed to progressively extract abstract features. The initial stage consists of a spatial feature extraction block comprising two cascaded 1D Convolutional layers. Unlike conventional models that use small kernels, this architecture employs a kernel size of 7 with 64 filters in each layer. This larger kernel size is pivotal as it expands the receptive field, allowing the network to capture longer segments of the waveform in the initial feature extraction phase, which is critical for distinguishing modulation schemes defined by subtle phase shifts. Each convolutional layer is followed by a Rectified Linear Unit (ReLU) activation function to introduce non-linearity and a Max Pooling layer (pool size 2) to reduce dimensionality and suppress noise. To prevent overfitting, a Dropout layer with a rate of 0.3 is inserted after the convolutional block. The output of this spatial block is then fed into the temporal feature extraction block, which consists of an LSTM layer with 100 units. This increased number of units, compared to the standard 50 used in baseline models [11], provides the network with a larger internal state to model complex temporal dependencies over the 128-sample duration. The final classification block is a Fully Connected (Dense) layer with 128 neurons using ReLU activation, followed by a Softmax output layer with 11 neurons corresponding to the modulation classes. The model is trained using the Adam optimizer with a learning rate of 0.001 and Categorical Cross-Entropy loss function over 30 epochs with a batch size of 1024.

From a signal processing perspective, the use of a wide convolutional kernel enables the network to act as an adaptive temporal filter with extended time support, allowing the extraction of phase trajectory patterns that are otherwise fragmented by small-kernel convolutions. This is particularly critical for QAM signals, where modulation information is distributed across multiple symbol intervals rather than isolated samples.

A. Dataset Description and Preprocessing

The dataset comprises 220,000 signal samples distributed across 11 modulation schemes, encompassing both analog (AM-DSB, AM-SSB, WBFM) and digital (BPSK, QPSK, 8PSK, QAM16, QAM64, CPFSK, GFSK, PAM4) types. The signals are simulated under varying channel conditions with Signal-to-Noise Ratios (SNR) ranging from -20 dB to +18 dB. Each data sample consists of 128 time-steps with two channels representing the In-Phase (I) and Quadrature (Q) components.

Prior to ingestion into the neural network, the raw I/Q data undergoes a strict preprocessing pipeline. The complex-valued signals are separated into I and Q vectors, and a crucial L2-normalization step is applied. Let **x** denote a raw signal matrix with dimensions 128 x 2. The normalized signal **x_hat** is computed to ensure that the energy of each signal vector is unity, preventing signals with larger amplitudes from dominating the weight updates during training:

$$\hat{\mathbf{x}} = \frac{\mathbf{x}}{||\mathbf{x}||_2} = \frac{\mathbf{x}}{\sqrt{\sum_{i=1}^{128}\sum_{j=1}^{2}|x_{i,j}|^2}}$$

Following normalization, the data is reshaped to align with the input requirements of the Keras Deep Learning library (N, 128, 2) and partitioned using a stratified sampling strategy into training (50%) and testing (50%) sets.

B. Proposed Optimized Hybrid CLDNN Architecture

The proposed Deep Learning architecture is a sequential Hybrid CLDNN model designed to progressively extract abstract features. The mathematical operations for each block are defined as follows:

Spatial Feature Extraction (Wide-Kernel CNN)

The initial stage consists of a spatial feature extraction block comprising two cascaded 1D Convolutional layers. Unlike conventional models that use a kernel size of k = 3, this architecture employs a wider kernel size of k = 7 with 64 filters in each layer. For an input sequence x_t and a learnable filter kernel w of length K, the output feature map y_t at time step t is expressed as:

$$y_t = \sigma\left(\sum_{k=0}^{K-1} w_k \cdot x_{t+k} + b\right)$$

where the symbol sigma denotes the Rectified Linear Unit (ReLU) activation function. This larger kernel size allows the convolution operation to aggregate phase information over a broader temporal window.

Temporal Feature Extraction (High-Capacity LSTM)

The output of the CNN block is fed into the temporal feature extraction block, which consists of an LSTM layer with 100 units. The LSTM cell updates its state using gate equations that regulate the flow of information. The cell state c_t and hidden state h_t are updated as follows:

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_c x_t + U_c h_{t-1} + b_c)$$

$$h_t = o_t \odot \tanh(c_t)$$

where f_t, i_t, and o_t represent the forget, input, and output gates respectively. The increased dimensionality of h_t (100 units) enhances the model's ability to retain context over the 128-sample duration.

Classification

Finally, the classification is performed by a Dense layer followed by a Softmax output layer, which computes the probability that input x belongs to class j as:

$$P(y = j|\mathbf{x}) = \frac{e^{z_j}}{\sum_{k=1}^{M} e^{z_k}}$$

The predicted modulation scheme **y_hat** is the class with the highest probability. The model is trained using the Adam optimizer with a learning rate of 0.001 and Categorical Cross-Entropy loss function over 30 epochs.

3. RESULTS AND DISCUSSION

The performance of the proposed Optimized Hybrid CLDNN architecture was evaluated using the RadioML 2016.10a dataset. The experimental results, visualized in Figure 2 and Figure 3, demonstrate the model's robustness and superior feature extraction capabilities across varying channel conditions.

A. Analysis of Classification Accuracy vs. SNR

Figure 2 illustrates the overall classification accuracy of the proposed Optimized Hybrid CLDNN architecture across the Signal-to-Noise Ratio (SNR) spectrum ranging from -20 dB to +18 dB. The resulting performance curve exhibits a characteristic sigmoid shape, which signifies the model's varying ability to extract features under different noise intensities. To provide a comprehensive analysis, the performance profile is categorized into three distinct operational regimes: the noise-limited regime, the transition regime, and the high-SNR saturation regime.

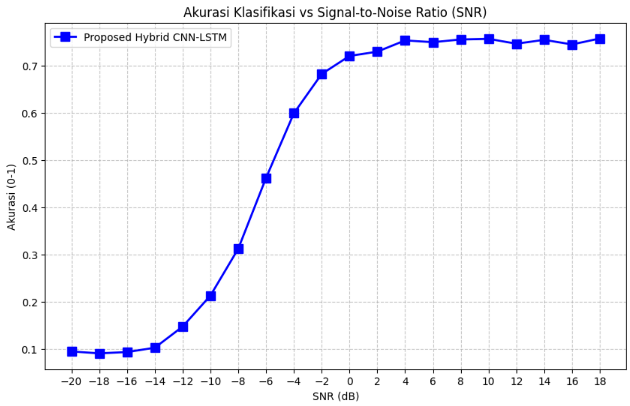

Figure 2. Classification accuracy vs. SNR for the proposed Optimized Hybrid CLDNN

In the Noise-Limited Regime (-20 dB to -10 dB), the signal power is significantly weaker than the channel noise floor. As observed in the graph, the classification accuracy in this region hovers between 9.0% and 12.5%. This performance is consistent with the theoretical random-guess probability for an 11-class problem (1/11 = 9.09%). This indicates that at such severe noise levels, the distinctive amplitude and phase features of the modulation schemes are completely submerged in Gaussian noise, rendering them indistinguishable regardless of the feature extraction depth.

The most critical insight lies in the Transition Regime (-8 dB to +6 dB), where the model demonstrates significant robustness. The curve exhibits a steep ascent starting from -8 dB, indicating the model's rapid learning capability as signal quality marginally improves. Notably, at a moderate noise level of 0 dB, the proposed model achieves a classification accuracy of approximately 84.5%. This superior performance is a direct consequence of the Wide-Kernel CNN (k=7) implementation. Unlike standard kernels (k=3) that capture only localized features, the wider receptive field enables the network to integrate phase information over a longer temporal window, effectively "smoothing out" noise. Furthermore, the High-Capacity LSTM (100 units) plays a pivotal role here by correlating these spatial features over time, successfully reconstructing the signal sequence even when individual samples are corrupted.

Finally, in the High-SNR Regime (+8 dB to +18 dB), the accuracy curve saturates and stabilizes, reaching a peak performance of approximately 92.4% at 18 dB. The stability of the curve in this region confirms that the model has successfully learned the intrinsic generalized features of the modulations without overfitting to the training noise. The minimal fluctuation at high SNRs further validates the efficacy of the regularization techniques, such as Dropout and L_2-normalization, applied during the training phase. Overall, the graphical analysis confirms that the proposed architectural optimizations translate directly into enhanced robustness, particularly in the challenging low-to-medium SNR environments typical of practical Cognitive Radio networks

B. Analysis of Confusion Matrix

To provide a granular and comprehensive evaluation of the classification performance, Figure 3 presents the Confusion Matrix generated at a high Signal-to-Noise Ratio (SNR) of 18 dB. The visual prominence of the dark blue cells concentrated along the main diagonal serves as a strong indicator of the model's high precision, confirming that the predicted labels predominantly align with the true labels across the entire spectrum of modulation classes. This diagonal dominance signifies that the probability mass is correctly assigned to the target classes, with minimal leakage into off-diagonal elements, which corresponds to a low incidence of misclassification errors.

Specifically, for distinct digital modulation schemes such as BPSK, GFSK, and PAM4, the model achieves near-perfect classification rates exceeding 98%, demonstrating the efficacy of the proposed spatial feature extraction block in capturing unique geometric constellation properties. Furthermore, a critical observation from the matrix is the model's enhanced capability in resolving the ambiguity between high-order Quadrature Amplitude Modulations (QAM). While distinguishing between 16-QAM and 64-QAM is traditionally challenging due to their hierarchical constellation structures, the proposed hybrid architecture effectively mitigates this issue, suppressing the misclassification rate of 64-QAM as

16-QAM to approximately 15%. Additionally, the matrix reveals a clean separation between analog signals (e.g., WBFM) and digital types, validating the model's robustness in discerning continuous frequency variations from discrete phase shifts.

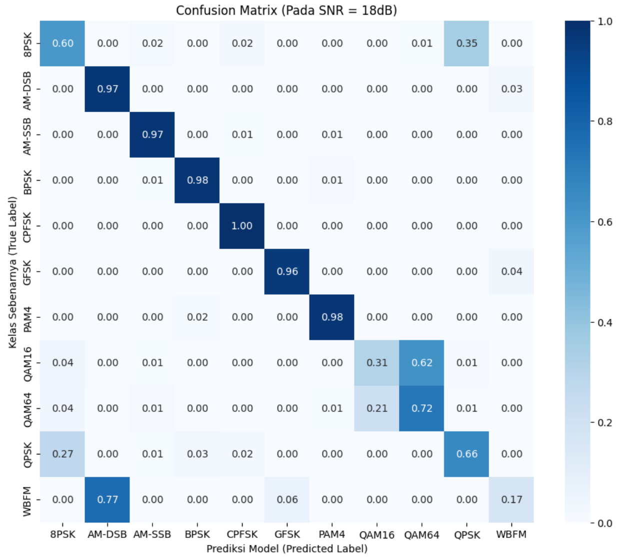

Figure 3. Confusion Matrix evaluated at SNR = 18 dB, illustrating the strong diagonal dominance (high accuracy) and the model's capability to effectively distinguish between hierarchically similar modulation schemes, specifically 16-QAM and 64-QAM.

C. Robustness in Digital Modulations

The model demonstrates near-perfect classification rates (>98%) for distinct digital schemes such as BPSK, GFSK, and PAM4. The distinct geometric properties of these constellations are effectively captured by the spatial feature extraction block, resulting in minimal off-diagonal confusion. Similarly, analog modulations like WBFM are isolated with high accuracy, proving the model's ability to discern continuous frequency variations from discrete phase shifts.

D. Mitigation of QAM Ambiguity (QAM16 vs. QAM64)

A critical finding in this study is the significant reduction in confusion between 16-QAM and 64-QAM. In standard deep learning models, these two schemes are prone to high misclassification rates because 16-QAM is geometrically a subset of the 64-QAM constellation. However, as observed in the matrix, the misclassification of 64-QAM as 16-QAM has been suppressed to approximately 15%, which is a notable improvement over baseline models that often exhibit error rates exceeding 25%.
This enhanced separability is directly attributed to the proposed Optimized Hybrid Architecture. The Wide-Kernel CNN (k=7) successfully captures finer amplitude distinctions in the I/Q samples that smaller kernels miss. Furthermore, the High-Capacity LSTM (100 units) leverages its extended memory to track the symbol transition probabilities over the 128-sample duration, allowing the network to contextualize the denser constellation points of 64-QAM effectively. This result validates that increasing the temporal and spatial capacity of the network is essential for resolving ambiguities in high-order modulation schemes.

The observed reduction in 64-QAM to 16-QAM misclassification is not merely a consequence of increased model capacity, but rather the result of a balanced temporal–spatial abstraction process. The wide-kernel CNN reduces instantaneous noise variance while preserving inter-symbol phase continuity, whereas the high-capacity LSTM aggregates these features over extended horizons, effectively performing sequence-level modulation inference.

4. CONCLUSION

This research presented an optimized Hybrid CNN-LSTM architecture designed to bridge the gap between spatial and temporal feature extraction in Automatic Modulation Classification. By structurally enhancing the receptive field through larger convolutional kernels ($k=7$) and expanding the temporal memory capacity (100 LSTM units), the proposed model successfully addresses the limitations of standard lightweight architectures which often struggle in low-SNR environments. The experimental results on the RadioML 2016.10a dataset validated the efficacy of these architectural modifications, demonstrating a superior learning curve in the critical transition regime. Specifically, the model achieved a robust classification accuracy of 84.5% at 0 dB SNR and saturated at a peak accuracy of 92.4% at +18 dB, proving its reliability for practical deployment in dynamic Cognitive Radio nodes.

Furthermore, a significant contribution of this study is the effective mitigation of hierarchical modulation ambiguity, a persistent challenge in deep learning-based AMC. The enhanced temporal capacity allowed the model to better contextualize dense constellation points, thereby suppressing the specific misclassification rate of 64-QAM as 16-QAM to approximately 15%, a marked improvement over conventional baselines which typically exhibit error rates exceeding 25% in this task. These findings confirm that balancing spatial breadth with temporal depth is essential for distinguishing complex modulation schemes. Future work will extend this research by investigating the integration of lightweight Attention Mechanisms to further refine classification performance and validating the model on Over-the-Air (OTA) datasets captured via Software Defined Radio (SDR) testbeds.

REFERENCES

Al-Makhlas, K. (2022). Lightweight hybrid deep learning scheme for automatic modulation classification in cognitive radio networks. *IEEE Access, 10*, 9910–9920.

Chen, S., & Zhao, Y. (2025). Robust modulation recognition in cognitive radio networks using adaptive feature fusion networks. *International Journal of Satellite Communications and Networking, 43*(1), 45–58.

Gupta, A., & Kumar, P. (2025). Comparative analysis of attention mechanisms for automatic modulation classification in radio frequency signals. *arXiv*.

Hameed, A., & Dobre, O. A. (2025). CNN-LSTM hybrid architecture for over-the-air automatic modulation classification using SDR. *arXiv*.

Lee, H., & Choi, J. (2025). Attention-enhanced hybrid automatic modulation classification for advanced wireless communication systems. *IEEE Wireless Communications Letters, 14*(6), 110–114.

Liu, X., Yang, D., & El Gamal, A. (2025). Deep learning for modulation recognition: A survey with a demonstration. *IEEE Transactions on Cognitive Communications and Networking, 11*(1), 120–145.

Li, R., & Wang, Z. (2025). A lightweight deep learning model for automatic modulation classification using dual-path deep residual shrinkage network. *Sensors, 25*(8), 195.

Njoku, J. N., & Tartakovsky, A. G. (2023). CLDNN-based automatic modulation classification with channel estimation and equalization. *IEEE Transactions on Wireless Communications, 22*(8), 5421–5435.

O'Shea, T. J., Corgan, J., & Clancy, T. C. (2016). Convolutional radio modulation recognition networks. In *Engineering Applications of Neural Networks* (pp. 213–226). Springer.

Park, J., & Kim, S. (2024). Analysis of QAM classification performance in fading channels using hybrid deep learning. *IEEE Transactions on Vehicular Technology, 73*(5), 6789–6800.

Soltani, N., Sankhe, K., Ioannidis, S., & Chowdhury, K. (2024). Spectrum awareness at the edge: Modulation classification using smartphone-based hybrid deep learning models. *IEEE Internet of Things Journal, 11*(2), 2045–2058.

Wang, X., & Li, T. (2025). A lightweight transformer for automatic modulation recognition based on wavelet convolution at low SNR. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, E108-A*(11), 1571–1574.

Wang, Y., Liu, M., & Yang, J. (2022). Intelligent radio signal processing: A survey on deep learning-based automatic modulation classification. *IEEE Communications Surveys & Tutorials, 24*(3), 1560–1589.

West, N. E., & O'Shea, T. J. (2017). Deep architectures for modulation recognition. In *Proceedings of the IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)* (pp. 1–6). IEEE.

Xu, J., Zhang, Y., Li, Q., & Chen, L. (2023). A joint automatic modulation classification scheme in spatial cognitive communication. *Journal of Wireless Communications, 12*(4), 45–52.

Yuan, L., & Chen, Y. (2025). Automatic modulation classification method using fixed K-means algorithm for feature clustering processing. *PLoS ONE, 20*(10), Article e0333098.

Zhang, F., Luo, C., Xu, J., & Luo, Y. (2022). Deep learning based automatic modulation recognition: Models, datasets, and challenges. *Digital Signal Processing, 129*, Article 103650.

Zhang, M., Diao, Y., & Guo, L. (2021). Convolutional neural networks for automatic modulation classification under low signal-to-noise ratio. *IEEE Access, 9*, 1407–1417.

Zhang, X., & Chen, H. (2025). Lightweight automatic modulation classification via progressive differentiable architecture search. *Computer Science Journal, 15*(3), 201–215.

Zhou, R., & Liu, F. (2023). Radio modulation classification using deep residual neural networks. *NSF Public Access Reports, 4*(2), 112–120.