Automatic Modulation Classification using Deep Learning

Pulijla Manjula
Entry No: 2024JTM2717
Telecommunication system and Management
Indian Institute of Technology, Delhi
New Delhi, India
jtm242717@dbst.iitd.ac.in

Jahnavi Daga
Entry No: 2024JTM2083
Telecommunication system and Management
Indian Institute of Technology, Delhi
New Delhi, India
jtm242083@dbst.iitd.ac.in

***Abstract*—An essential component of intelligent wireless communication systems, Automatic Modulation Classification (AMC) plays a crucial role in enabling adaptive demodulation and signal awareness. End-to-end modulation classification has advanced significantly with the advent of deep learning, especially convolutional neural networks (CNNs), which learn discriminative features directly from raw in-phase and quadrature (IQ) samples. In this study, we assess and contrast a new parallel CNN-transformer network (PCTNet) intended to enhance classification performance with a traditional CNN architecture. In order to efficiently model long-range dependencies over the signal, the suggested PCTNet blends transformer-based modules with CNNs' prowess in capturing local temporal patterns. With an integrated fusion mechanism to improve feature representation, the CNN and transformer routes function in parallel, in contrast to conventional cascaded architectures. According to experimental results, PCTNet performs better in classification accuracy than solo CNN models, indicating its potential for reliable and effective AMC in challenging wireless situations.**

***Index Terms*—Automatic modulation classification (AMC), convolutional neural networks (CNNs), transformer networks, deep learning, wireless communication.**

I. Introduction

AMC allows a receiver to independently determine the modulation scheme of an incoming signal without any prior information, acting as a transitional stage between signal detection and demodulation. Enhancing the intelligence, adaptability, and interoperability of communication systems functioning in dynamic and varied spectrum environments requires this capacity.

AMC methods have historically been divided into two categories: likelihood-based and feature-based. Even though likelihood-based approaches perform best in the best-case scenario, they are computationally costly and necessitate exact knowledge of channel properties. Conversely, feature-based approaches depend on expertly created signal characteristics like cyclic cumulants or higher-order statistics. However, when noise, multipath fading, or other channel impairments are present, these methods frequently see a decline in performance. Since deep learning is developing so quickly, data-driven methods have becoming strong substitutes for AMC. Because they can learn discriminative features straight from raw IQ samples, Convolutional Neural Networks (CNNs) in particular have showed promise by doing away with the requirement for manually created features. CNNs are excellent at identifying local patterns in data, but they frequently have trouble simulating global contextual linkages and long-range dependencies in the signal.

Recent research has looked into integrating transformer topologies, which are well-known for their ability to successfully describe sequential data via self-attention techniques, in order to get around this restriction. Transformers are good at capturing global dependencies, but they don't have CNNs' inductive biases, which are useful for simulating local signal structures.

Motivated by the aforementioned work, we construct and analyze a parallel CNN-transformer architecture for AMC in this research. The model employs a dual-path structure in which a fusion mechanism incorporates both local and global information after CNN and transformer modules separately extract features from raw IQ signals. We compare it against a CNN-based design. Our results demonstrate the complimentary nature of convolution and attention-based representations in AMC tasks by confirming that the suggested framework considerably improves classification accuracy across a range of signal-to-noise ratios.

II. Problem statement

In contemporary wireless communication systems, Automatic Modulation Classification (AMC) is essential because it acts as a link between signal detection and demodulation. Its uses are essential in fields including electronic warfare, cognitive radios, and spectrum monitoring. The two main categories of traditional AMC methods are feature-based and likelihood-based approaches. Although likelihood-based approaches can be the best in certain situations, they are computationally costly and require prior knowledge of the channel characteristics. The resilience and flexibility of feature-based approaches to changing noise and channel conditions are limited by their heavy reliance on handmade characteristics such as cyclostationary qualities and higher-order statistics.

Deep learning methods, especially Convolutional Neural Networks (CNNs), have been used to get beyond these restrictions. CNNs exhibit good performance over a range of signal-to-noise ratios (SNRs) and are capable of autonomously extracting discriminative features from raw In-phase and Quadrature (IQ) signals. However, because of their restricted receptive field, CNNs are intrinsically constrained in their ability to model long-range relationships. As a result, they are less successful in capturing global structures and temporal dynamics in modulation patterns.

Transformer architectures have been added to AMC in recent methods to solve this weakness. Originally created for natural language processing, transformers can be used to understand the temporal structure of communication signals since they simulate global dependencies via self-attention techniques. By combining CNNs with transformers either sequentially or concurrently, a new class of hybrid models seeks to take advantage of both the global context modeling of transformers and the local feature extraction capabilities of CNNs.

Motivated by previous research showing the higher classification performance of such hybrid architectures, we concentrate on creating and assessing a parallel CNN-transformer framework for AMC in this work. By efficiently capturing both local and global signal features, we want to develop a model that improves modulation classification accuracy without the need for manual feature engineering or domain-specific preprocessing such as FFT. The three main issues in AMC that this integrated approach tackles are learning discriminative representations from raw IQ inputs, generalizability, and robustness to noise.

III. Model Architecture

The architecture of the suggested Parallel CNN-Transformer Network (PCTNet), which is intended for automated modulation categorization (AMC) that is both effective and efficient, is described in depth in this section. The architecture is especially well-suited for processing complex and noisy wireless communication signals because it combines the global contextual modeling power of Transformer encoders with the local feature extraction skills of Convolutional Neural Networks (CNNs).

Raw in-phase (I) and quadrature (Q) baseband samples from received radio signals make up the model's input. A two-channel sequence is used to represent each input sample, with one channel representing the I component and the other the Q component. An orderly two-dimensional matrix of shapes is created from the I/Q data.

N stands for the number of sample points per signal. As input to the CNN and Transformer modules in parallel, this format is essential to maintain the temporal dynamics and structure of the modulated signal.

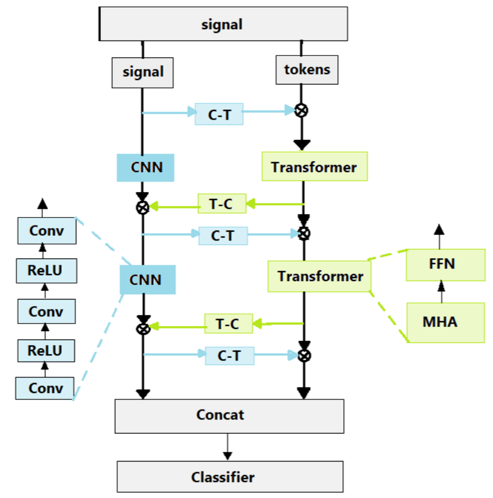


Fig. 1. PCTNet Architecture

A. Convolutional Feature Extractor

The local spatial properties of the input signal must be captured by the CNN module. Three convolutional layers make up this system, and its kernel size is 5X5 with the goal of gradually learning hierarchical representations from the unprocessed signal. A ReLU activation function, which adds non-linearity, and batch normalization, which stabilizes and speeds up training, come after each convolutional layer. In order to minimize spatial dimensions and allow the network to concentrate on the most important signal properties, the CNN block furthermore has max pooling layers. Crucially, the CNN architecture is based on residual networks, which permits linear channel growth in each layer and makes it easier to extract multiscale feature maps that are necessary for long-sequence representation.

B. Transformer Structure

PCTNet utilizes a transformer encoder layer but removes the standard residual connections. Instead, multi-head self-attention (MHSA) is directly followed by a feedforward network (FFN).

The MHSA mechanism is defined as:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h)W^O \tag{1}$$

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \tag{2}$$

The FFN is a two-layer MLP with a ReLU activation:

$$\text{FFN}(x) = \text{ReLU}(xW_1 + b_1)W_2 + b_2 \tag{3}$$

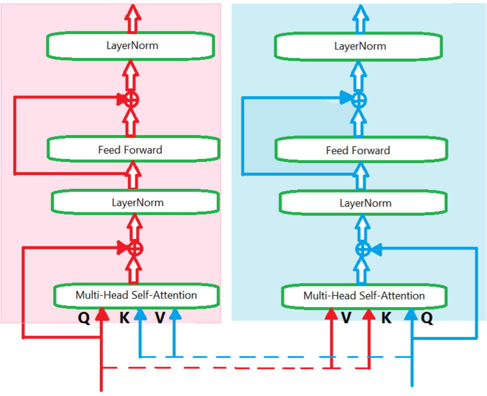


Fig. 2. Exchanging key and values in cross attention

C. CNN and Transformer Delivery Mechanisms

1) Cross Attention Mechanism: Inspired by multimodal fusion methods, the delivery mechanism adopts a cross-attention where keys and values are interchanged. Let:

$$Y = \begin{bmatrix} R \\ E \end{bmatrix}, \quad K_Y = YW^K = \begin{bmatrix} K_R \\ K_E \end{bmatrix}, \quad Q_Y = YW^Q = \begin{bmatrix} Q_R \\ Q_E \end{bmatrix},$$

$$V_Y = YW^V = \begin{bmatrix} V_R \\ V_E \end{bmatrix}$$

Then the scaled dot-product attention becomes:

$$\text{Attention}(Q_Y, K_Y, V_Y) = \text{softmax}\left(\frac{Q_Y K_Y^T}{\sqrt{d}}\right) V_Y \tag{4}$$

Which can be expanded into:

$$Q_Y K_Y^T V_Y = \begin{bmatrix} Q_R K_R^T V_R + Q_R K_E^T V_E \\ Q_E K_E^T V_E + Q_E K_R^T V_R \end{bmatrix} = \begin{bmatrix} R_{\text{up}} \\ E_{\text{up}} \end{bmatrix}$$

2) Directional Attention: For $C \to T$, the CNN features X are directly used as key and value without projection. The attention mechanism becomes:

$$X \to Z = \left[\text{Attention}(\tilde{z}_i W_i^Q, \tilde{x}_i, \tilde{x}_i)\right]_{i=1}^{h} W^O \tag{5}$$

For $T \to C$, transformer outputs Z are keys and values; CNN outputs are the queries:

$$Z \to X = \left[\text{Attention}(\tilde{x}_i, \tilde{z}_i W_i^K, \tilde{z}_i W_i^V)\right]_{i=1}^{h} \tag{6}$$

Here, $\tilde{x}_i$ and $\tilde{z}_i$ represent the i-th split head from X and Z respectively.

IV. Experimental Data and Analysis

A. Datasets

The model is evaluated in a synthetic dataset that is generated by us, which simulates real wireless environments, including fading, noise.Dataset parameters as follows: Total samples: 12,000, Modulation types: BPSK, QPSK, 16-QAM, Noise: AWGN, N=512, Fading: Rayleigh, SNR=-5 to 20dB

B. Training Parameters

The network is trained using the following hyperparameters: learning rate = 0.0004, batch size = 128, decay rate = 0.5, dropout = 0.2, weight decay = 1e-5.

C. Results

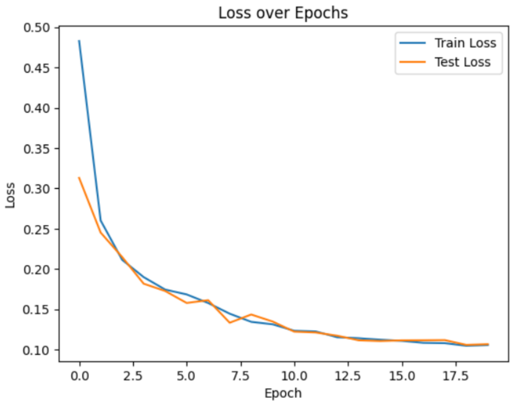


Fig. 3. CNN-Transformer Train/Test Loss vs. Epochs

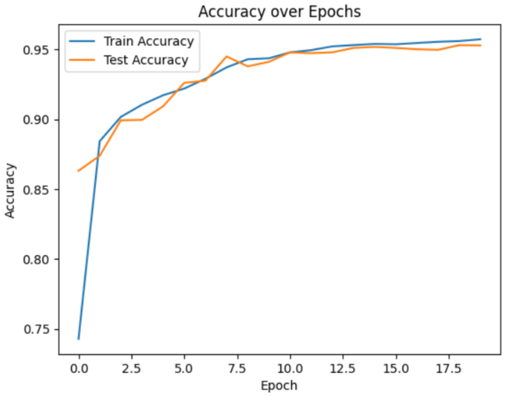
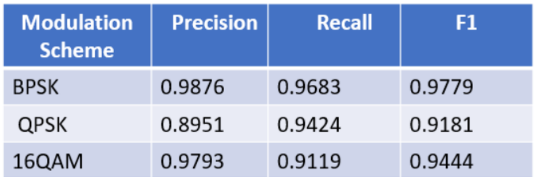


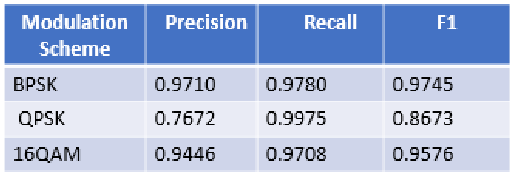
Fig. 4. CNN-Transformer Accuracy vs. Epochs



Modulation Scheme	Precision	Recall	F1
BPSK	0.9876	0.9683	0.9779
QPSK	0.8951	0.9424	0.9181
16QAM	0.9793	0.9119	0.9444

Fig. 5. CNN-Transformer Metrics

The results of fig. 3, 4 highlight the efficiency and reliability of the proposed parallel CNN-Transformer model for automatic modulation classification. The model demonstrates fast convergence, attaining over 90% accuracy within the initial five epochs and achieving a maximum test accuracy of 95.31% by the 19th epoch. During training, the loss and accuracy metrics steadily improve, with no indication of overfitting, as test performance closely mirrors the training



Modulation Scheme	Precision	Recall	F1
BPSK	0.9710	0.9780	0.9745
QPSK	0.7672	0.9975	0.8673
16QAM	0.9446	0.9708	0.9576

Fig. 6. CNN Metrics

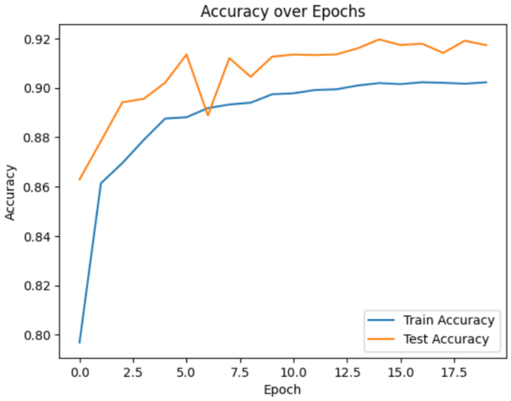


Fig. 7. CNN Train/Test Accuracy vs. Epochs

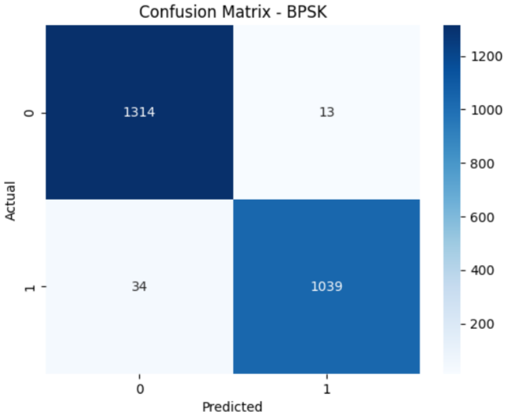


Fig. 8. CNN Transformer BPSK confusion matrix

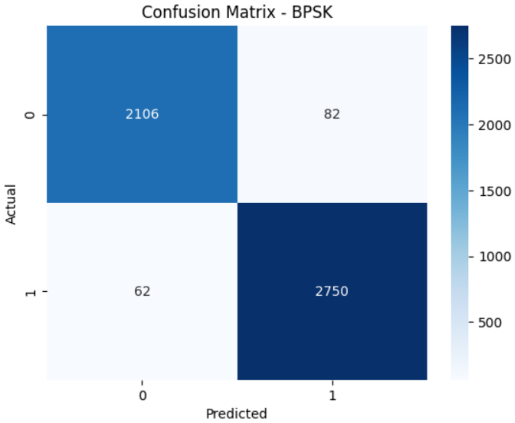


Fig. 9. CNN BPSK confusion matrix

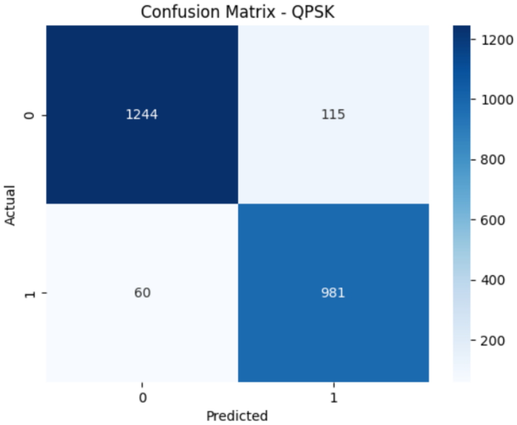


Fig. 10. CNN Transformer QPSK confusion matrix

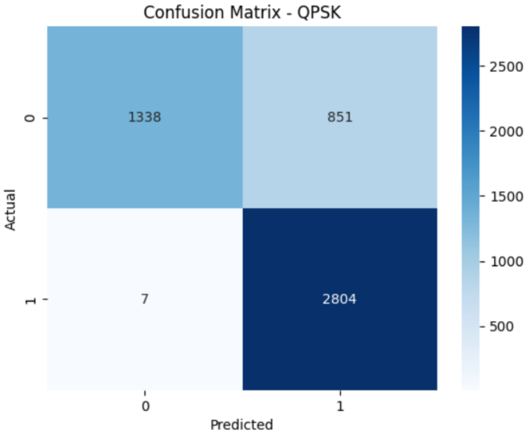


Fig. 11. CNN QPSK confusion matrix

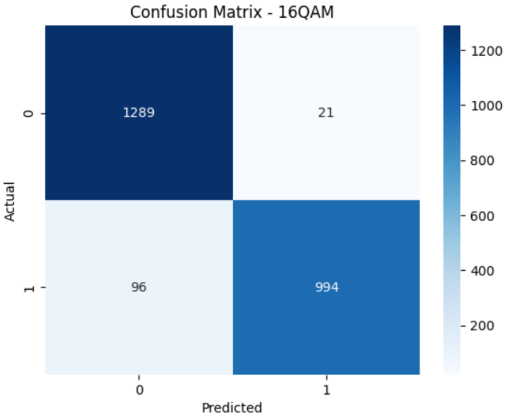


Fig. 12. CNN Transformer 16-QAM confusion Matrix

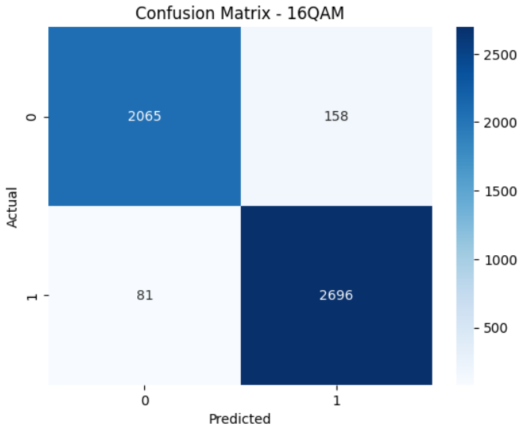


Fig. 13. CNN 16QAM confusion matrix

patterns. The consistent reduction in loss values combined with the stabilization of accuracy suggests that the model successfully captures both local and global features essential for modulation type classification. From fig. 5 depicts that the metrics for each class and the overall detection scores highlight the model's robustness and impressive accuracy in differentiating various modulation types. BPSK exhibited the best overall performance, achieving a precision of 98.76% and an F1 score of 97.79%, which indicates a low rate of misclassification. QPSK and 16QAM also performed well, recording F1 scores of 91.81% and 94.44% respectively, showcasing the model's capability to generalize effectively across modulation schemes of differing complexities. The overall detection metrics are exceptionally high, with flawless precision and almost perfect recall (99.69%), resulting in an exceptional F1 score of 99.84%. These findings, corroborated by the confusion matrix, affirm that the model excels at distinguishing between modulation classes, demonstrating its applicability for real-world AMC scenarios.

From Figs.8, 10 and 12 confusion matrices for BPSK, QPSK, and 16QAM modulation schemes, it is clear that the model exhibits strong classification capabilities with high true positive rates for all classes. In the case of BPSK, the majority of samples are accurately classified, with only a few misclassifications—13 false positives and 34 false negatives—highlighting solid precision and recall. Likewise, for QPSK, while there are slightly more misclassifications (115 false positives and 60 false negatives), most predictions remain accurate, reflecting consistent performance. For 16QAM, the model also correctly classifies the majority of instances, although it shows a somewhat higher number of false negatives (96) compared to false positives (21). Overall, these confusion matrices confirm that the model upholds reliable prediction accuracy across various modulation types, reinforcing high F1 scores and precision metrics.

V. Conclusion

Based on results,The CNN Transformer model outperformed CNN with a smaller test loss of 0.12 vs. 0.20 and a higher test accuracy of 95.6% vs. 91.7%. Eventhough both models performed well on BPSK, CNN struggled with QPSK, misclassified 851 samples, compared to the hybrid model's 115 misclassifications. Because of its global attention mechanism the Transformer was able to recognise complex patterns in modulations like QPSK and 16QAM. As a result, QPSK's F1 increased from 0.8673 to 0.9173, and precision and F1-scores improved for all classes. All things considered, the CNN-Transformer demonstrated more consistent learning and performed better across all signal types.

VI. References

1. W. Ma, Z. Cai and C. Wang, ”A Transformer and Convolution-Based Learning Framework for Automatic Modulation Classification,” in IEEE Communications Letters, vol. 28, no. 6, pp. 1392-1396, June 2024, doi: 10.1109/LCOMM.2024.3380623.
2. H. Xing et al., ”Joint Signal Detection and Automatic Modulation Classification via Deep Learning,” in IEEE Transactions on Wireless Communications, vol. 23, no. 11, pp. 17129-17142, Nov. 2024, doi: 10.1109/TWC.2024.3450972.
3. C. Hou, G. Liu, Q. Tian, Z. Zhou, L. Hua and Y. Lin, ”Multisignal Modulation Classification Using Sliding Window Detection and Complex Convolutional Network in Frequency Domain”in IEEE Internet of Things Journal, vol. 9, no. 19, pp. 19438-19449, 1 Oct.1, 2022, doi: 10.1109/JIOT.2022.3167107.