

## RESEARCH ARTICLE

# A Novel Approach for Multicarrier Modulation Signals Using GTCC Spectrogram and Improved Transfer Learning Models

Km. Sejal | Mohit Dua 

Department of Computer Engineering, National Institute of Technology, Kurukshetra, India

**Correspondence:** Mohit Dua ([er.mohitdua@nitkkr.ac.in](mailto:er.mohitdua@nitkkr.ac.in))

**Received:** 27 February 2025 | **Revised:** 17 June 2025 | **Accepted:** 4 July 2025

**Funding:** The authors received no specific funding for this work.

**Keywords:** automatic modulation classification | GTCC | MCM | SNR | spectrograms

## ABSTRACT

High-speed and low-latency communication are critical for next-generation wireless networks, demanding accurate and computationally efficient signal demodulation. However, conventional demodulation techniques suffer from high complexity and poor robustness under low signal-to-noise ratio (SNR) conditions, limiting real-time applicability. This study presents a novel AMC framework for 5G systems using simulated multicarrier modulation (MCM) signals from the Vienna 5G Link Level Simulator with Jake's Doppler model and varying SNRs via added Gaussian noise. The front end of the proposed method uses an innovative feature extraction approach using gammatone cepstral coefficient (GTCC) spectrograms that capture fine-grained spectral-temporal patterns of MCM signals, enhancing feature discrimination. At the back end of the proposed scheme, three enhanced deep learning models, Improved-InceptionV3, Improved-Xception, and Improved-ResNet152V2, have been applied, individually, integrating fine-tuned convolutional layers, dense blocks, dropout regularization, and adaptive learning rates for superior performance. Experiments on the 5G waveform dataset demonstrate state-of-the-art classification accuracy, with Improved-ResNet152V2 leveraging architectural pruning and quantization outperforming Improved-InceptionV3 and Improved-Xception, achieving 96% at  $-16\text{ dB}$  SNR and 100% at  $16\text{ dB}$  SNR. The synergy of GTCC spectrograms and enhanced architectures enables robust accuracy-efficiency trade-offs, outperforming existing methods in low SNR and noisy conditions, while Grad-CAM visualizations enhance interpretability and reliability for AMC tasks. This framework establishes a significant advancement in AMC research.

## 1 | Introduction

The 5G generation has raised expectations for better dependability, reduced latency, and quicker data speeds. Ensuring the fulfillment of this requirement calls for a communication infrastructure that is robust, safe, and effective. The fundamental technology known as modulation makes it possible for cell phones to seamlessly exchange radio waves and for data to be transmitted quickly across fiber optic cables. As a versatile foundation of communication, modulation reduces latency for fast,

almost real-time responsiveness, uses high-frequency carriers to ensure stable transmission over long distances, and permits greater data rates through signal compression. Additionally, by adding intentional noise to signals, preventing unlawful interception and eavesdropping efforts, and safeguarding data privacy, several modulation techniques improve data security.

Many different industries and domains use automatic modulation classification (AMC), wireless communication networks being one of the most prominent examples. Within this domain, AMC

is key in classifying the modulation schemes of incoming signals, which in turn optimizes resource allocation, signal processing, and decoding. Through this optimization, system performance is improved, and performance quality is raised across changing communication protocols, including 4G, 5G, and beyond.

Multicarrier modulation (MCM) signals are a type of modulation technique where multiple carrier frequencies are used to transmit data simultaneously. Orthogonal frequency division multiplexing (OFDM) is the most often used type of MCM. It splits a single high-speed data stream into multiple slower data streams that are simultaneously sent on various frequencies. This approach significantly enhances data transmission efficiency and robustness, making it ideal for environments with high levels of interference or multipath propagation, such as urban areas or indoor settings. MCM signals play a crucial role in modern communication systems, especially in 5G technologies, which require low latency, high data rates, and the ability to handle a large number of devices. The flexibility and adaptability of MCM make it suitable for diverse applications, including mobile broadband, machine-to-machine (M2M) communications, and the Internet of Things (IoT). Given the importance of MCM in 5G networks, the need for efficient and accurate modulation classification techniques has become increasingly important to ensure optimal performance and resource allocation.

Several studies have explored the development of effective automatic AMC techniques, which are essential for cognitive radio and dynamic spectrum access networks. Traditional AMC methods typically rely on feature extraction based on the signal's statistical properties, cyclostationary features, or higher order cumulants, but these approaches often fail to generalize across different communication environments or modulation schemes. Recent advancements in machine learning, particularly deep learning, have introduced more robust solutions by leveraging large datasets and complex models for higher classification accuracy. Among these advancements, gammatone cepstral coefficients (GTCCs) have emerged as a promising feature extraction technique due to their unique characteristics. Derived from the gammatone filter bank, which mimics the frequency response of the human auditory system, GTCCs effectively capture subtle variations in the spectral content of signals, making them highly suitable for distinguishing between different modulation types, particularly in scenarios where traditional methods fall short. GTCCs offer enhanced spectral resolution compared to conventional features like mel-frequency cepstral coefficients (MFCCs) or short-time Fourier transform (STFT), allowing for a more detailed analysis of a signal's frequency components and improving classification accuracy in noisy or interference-heavy environments. Additionally, GTCCs are robust to noise and interference common challenges in real-world communication systems by emphasizing spectral peaks and reducing the impact of background noise, thereby maintaining high classification accuracy even under adverse conditions often encountered in 5G networks. Their effectiveness in representing nonstationary signals an essential requirement for AMC given the time-varying nature of modulation signals enhances AMC algorithms' performance by providing a more accurate depiction of a signal's spectral behavior over time.

Recent developments in deep learning and transfer learning have led to the introduction of improved AMC models, such as Improved-InceptionV3, Improved-Xception, and Improved-ResNet152V2. ImageNet pretrained architectures are leveraged by replacing the original classification heads with a global average pooling (GAP) layer to condense spatial-spectral features from GTCC spectrograms, followed by task-specific dense layers with ReLU activation and a softmax output layer reconfigured for 10 modulation classes. To bridge the domain gap between natural images and GTCC spectrograms of modulating signals, a two-phase training strategy is employed. During the initial phase, base model weights are frozen to preserve hierarchical filters (e.g., edge and texture detectors) that inherently capture cyclostationary patterns and spectral correlations in low signal-to-noise ratio (SNR) environments, allowing optimization to focus solely on mapping these generalized representations to modulation schemes. In the subsequent phase, the entire network is unfrozen and fine-tuned with a reduced learning rate, enabling noise-adaptive recalibration of high-level features (e.g., constellation geometry and phase transitions) critical for discriminating subtle modulation differences, while gradient clipping and layer-wise learning rate decay mitigate catastrophic forgetting. This approach enhances low SNR robustness and ensures real-time applicability by aligning pretrained feature hierarchies with domain-specific discriminators, making it viable for deployment in resource-constrained 5G edge devices. Among these, Improved-ResNet152V2 demonstrates the highest classification accuracy, achieving 96% accuracy at -16dB and 100% accuracy at 16dB SNR, making it highly suitable for real-world communication environments with varying noise conditions.

By integrating these improved transfer learning models with GTCC features, the proposed AMC framework surpasses traditional convolutional neural network (CNN)-based approaches in both accuracy and computational efficiency. This technique ensures robust performance in dynamic and noisy environments, addressing the challenges of modern wireless communication systems. The advancements presented in this study contribute significantly to enhancing the reliability and security of 5G networks, supporting future developments in high-speed, low-latency communication technologies. The significant contributions of the proposed work can be described as follows: A front-end framework is proposed for extracting discriminative spatial-spectral features from MCM signals using GTCC spectrograms, enhancing noise-robust representation in dynamic channels, which is novel to MCM and AMC. Three enhanced ImageNet pretrained models, Improved-InceptionV3, Improved-Xception, and Improved-ResNet152V2, are integrated with GTCC spectrograms for classification, leveraging architectural refinements to optimize computational efficiency and accuracy.

The remaining sections of this paper are structured as follows: Section 2 presents a review of previous research on AMC techniques. Section 3 describes the methodologies used to implement the proposed approach. Section 4 provides an overview of the proposed system. Section 5 presents experimental results, followed by a discussion and comparative analysis in Section 6, and Section 7 concludes the study.

## 2 | Related Work

Extensive research has optimized both feature extraction and classification techniques for modulated signals. Ho et al. [1] focused on BPSK (binary phase shift keying) and BASK (binary amplitude shift keying) modulation types at  $-5\text{ dB}$  SNR, achieving 54% accuracy with wavelet transform for feature extraction. Xie et al. [2] used cumulants for feature extraction from six modulation types, reaching 99% accuracy at  $-5\text{ dB}$  and 100% at  $-2\text{ dB}$  SNR. Moser et al. [3] applied instantaneous features for classifying nine modulation types, attaining 100% accuracy at SNRs of 10 and 20 dB with a threshold of 8192. Peng et al. utilized constellation diagrams to achieve 100% accuracy in classifying BPSK, 4ASK (amplitude shift keying), QPSK (quadrature phase shift keying), OQPSK (offset quadrature phase shift keying), and 8PSK (phase shift keying) at 4 dB SNR. Wu et al. [4] used STFT to obtain features from 16 modulation types, achieving 99% accuracy by converting features into grayscale spectrograms. Tailor et al. [5] also employed STFT for generating spectrograms of MCM signals. Jdid et al. [6] adapted MFCCs and GTCCs for modulation classification, capturing perceptually relevant aspects of signals and enhancing performance. GTCC spectrograms were selected for their robustness to noise and their ability to integrate both time and frequency information.

For classification, CNNs demonstrated effectiveness in analyzing modulation schemes by learning spatial hierarchies in transformed signal representations. O'Shea et al. [7] highlighted CNNs' robustness to signal variations, making them suitable for AMC. Long short-term memory (LSTM) networks were employed to capture time-dependent modulation patterns, as shown by Rajendran et al. [8]. Hu et al. compared deep learning-based recurrent neural networks (RNNs) with other models, while Duan et al. [9] reviewed machine learning-based AMC approaches for MIMO and SISO systems. Wu et al. compared CNNs with support vector machines (SVMs) for classifying 16 modulation signals, achieving 99% accuracy with CNNs compared to 78% with SVMs. Recent advances in AMC emphasize lightweight architectures and hybrid deep learning frameworks to address computational efficiency and noise robustness. Guo et al. [10] proposed an ultralite convolutional neural network (UL-CNN) for AMC in UAV-IoT networks, achieving real-time classification with minimal parameters. However, their work

focuses on narrowband signals and does not address MCM complexities. Ma et al. [11] introduced a transformer-convolution hybrid framework for AMC, demonstrating superior accuracy over ResNet on the RadioML2018 dataset. While their model leverages self-attention for global feature extraction, its computational overhead limits deployability in resource-constrained 5G edge nodes. Ouamna et al. [12] explored spectrogram-based AMC using CNNs on the RadioML2016.10a dataset, reporting 92% accuracy for BPSK and QPSK classification. Their comparison with VGG16, ResNet18, and AlexNet highlights the trade-off between model depth and inference latency. However, their work is restricted to single-carrier modulations and does not evaluate performance under extreme SNR conditions. These studies collectively highlight the effectiveness of integrating advanced feature extraction techniques with deep learning models for improved modulation signal classification, as shown in Table 1.

The related literature has been examined, and it is evident that deep learning methods, in particular, CNNs and RNNs [14–16], have consistently demonstrated their effectiveness in automating a variety of classification tasks, most notably modulation classification. Tailor et al. [5] used CNN variants, including CNN, CNN with both dropout and L1 regularization (CNN\_dL1), CNN with dropout (CNN\_d), and the Xception model to classify STFT spectrograms, and out of these, Xception performs well with 98% accuracy. Motivated by this research, the idea of this research paper is to extract the features from the MCM signals using the GTCC spectrograms at the frontend. Improved transfer learning models such as InceptionV3, Xception, and ResNet152V2 are utilized for the classification of the signals at the backend. The following is a summary of this paper's contribution:

1. This paper proposes a front-end framework that uses GTCC spectrograms for the extraction of dominant features from the given MCM signal.
2. Three improved transfer learning models—InceptionV3, Xception, and ResNet152V2 are employed with GTCC spectrograms for the classification of MCM signals in the backend.

**TABLE 1** | Comparison with existing approaches.

Existing work	Key contribution	Limitations	Proposed model's improvements
Ma et al. [11]	Hybrid CNN transformer (PCTNet) achieves more than 90% accuracy on 0 dB and 93.45% accuracy on 10 dB SNR.	High computational overhead untested below 0 dB SNR.	Improved transfer learning models applied on less than 0 dB SNR
Molis et al. [13]	STFT spectrograms for feature extraction and transfer learning model Xception for classification applied on $-10$ to $16\text{ dB}$ SNR	STFT lacks noise resilience at ultralow SNR ( $-16\text{ dB}$ ).	GTCC's auditory-inspired features achieve 96% at $-16\text{ dB}$ , outperforming STFT-based Xception.
Ouamna et al. [12]	Spectrogram-based CNN achieves 92% accuracy for BPSK/QPSK on RadioML2016.10a.	Evaluates only single-carrier modulations; no MCM or low SNR analysis.	10-class MCM taxonomy (5 waveform types $\times$ 2 QAM variants) tested at $-16\text{ dB}$ to $+16\text{ dB}$ SNR.

3. There are precisely 10 different MCM signals in the dataset that are used for classification. There are five different MCM signal types taken into consideration: OFDM, windowed orthogonal frequency division multiplexing (WOLA), filtered orthogonal frequency division multiplexing (FOFDM), filter bank multicarrier (FBMC), and universal filtered multicarrier (UFMC). Additionally, two subcarrier waveform types, QAM16 and QAM64, for every signal have been considered, resulting in a total of 10 distinct MCM signals that can be classified.
4. To evaluate the noise robustness of the proposed framework, tests are conducted across an SNR range from  $-16$  to  $+16$  dB in 4 dB increments. Performance is assessed using metrics such as accuracy, precision, recall, and  $F_1$  score. Furthermore, the proposed models are evaluated on inputs with added colored noise, and Grad-CAM visualizations are employed to analyze feature attribution and model interpretability.

### 3 | Preliminaries

In this section of the paper, we delve into the technical aspects that underpin our work, focusing specifically on the intricate processes involved in transfer learning models and the generation of spectrograms for communication signals using GTCC.

#### 3.1 | GTCC Spectrogram

Spectrograms are crucial for analyzing and interpreting communication signals. They visually represent the frequency content of a signal over time. In the field of communication signals, including audio, speech, and radio frequency transmissions, spectrograms are vital for examining the detailed features of these signals. By transforming time-domain signals into the frequency domain, spectrograms allow for the identification of signal components, modulation patterns, and the detection of any anomalies or interference. GTCC is mainly used in speech recognition and language recognition to extract features that are useful to classify them [17, 18]. GTCC is used to classify MCM signals because they are the 5G signals used for communication through mobile phones. Basically, those are the speech or voice signals only. Our main motive for using this technique is its noise robustness.

The name “gammatone” comes from the fact that the filter’s impulse response is the product of a gamma distribution and a sinusoidal tone. Mathematically, the impulse response  $g(t)$  of a gammatone filter is given by:

$$g(t) = t^{n-1} e^{-2\pi bt} \cos(2\pi f_c t + \varphi) \quad (1)$$

where  $t$  is the time,  $n$  is the filter order,  $b$  is the bandwidth,  $f_c$  is the center frequency of the filter, and  $\varphi$  is the phase.

The GTCC depends on the gammatone filter bank. First, the speech signal undergoes pre-emphasis to enhance its quality [19]. Subsequently, the signal is segmented into frames ranging from 5 to 10 ms. To eliminate spectral distortion, each frame is windowed. The windowed signal is then processed using FFT (fast Fourier transform) to obtain the FFT spectrum. Following this, a gammatone filter bank is applied to simulate the frequency selectivity characteristic of the human cochlea. Figure 1 states the steps involved in GTCC spectrogram generation:

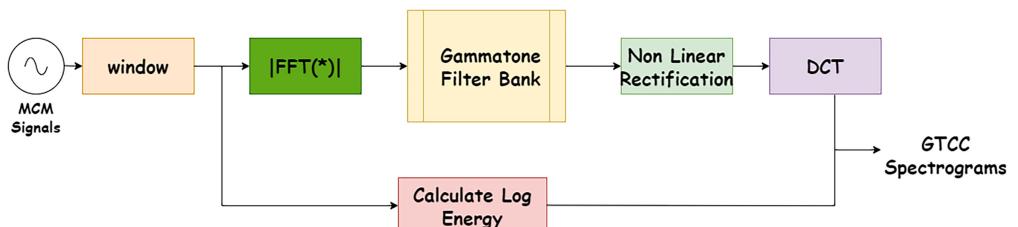
Generating a GTCC spectrogram involves several key steps. First, apply pre-emphasis to the signal to enhance high frequencies. Next, divide the signal into overlapping frames and apply a Hamming window to each frame. Then, pass each frame through a gammatone filter bank to simulate auditory filtering. Calculate the energy of each filter’s output and apply a logarithmic transformation. Apply the discrete cosine transform (DCT) to these log energies to obtain the GTCCs. Finally, stack the GTCC vectors from each frame to create the 2D GTCC spectrogram.

#### 3.2 | Transfer Learning-Based Deep Learning Models

A model created for one task is utilized as the foundation for a new model for another task, a process known as transfer learning in machine learning. This method is especially helpful in situations when there is a lack of labeled data since it enables models to make use of information that has already been gathered from sizable datasets for fundamental tasks like ImageNet’s image classification. With minimum additional training, transfer learning allows for faster convergence and better performance on target tasks by optimizing pretrained models such as Inception [20, 21], Xception [22], and ResNet152V2 [10, 23].

AMC relies on deep learning models for robust signal classification [24, 25]. Architectural enhancements such as additional dense layers [26], optimized hyperparameters, and fine-tuning techniques improve accuracy, generalization, and robustness under varying SNRs.

- *Additional dense layers for feature representation:* Dense layers perform nonlinear transformations on extracted features:



**FIGURE 1** | GTCC spectrogram generation process.

- *Optimized hyperparameters:* Hyperparameters significantly impact a model's learning process and classification performance. By carefully tuning these hyperparameters, transfer learning models achieve improved modulation recognition, adapting to variations in real-world wireless environments. The optimized hyperparameters include learning rate, batch size, and regularization techniques, which contribute to building an improved model.

- Efficient learning rates* that prevent slow convergence or overshooting of optimal weights. A lower learning rate during fine-tuning ensures stable gradient updates, refining pretrained weights without distorting prior knowledge as given in Equation (3).
- Batch size adjustments* that balance computational efficiency with gradient stability, preventing overfitting or underfitting.
- Regularization techniques* such as dropout and batch normalization, which reduce overfitting while maintaining computational efficiency.
- Fine-tuning adapts pretrained models to modulation-specific spectrograms by refining high-level feature extraction in deeper layers. This enhances classification robustness across varying SNR conditions while leveraging learned representations for improved accuracy. By adjusting filters to wireless communication signals, the model retains general features while focusing on task-specific patterns. This approach ensures efficient transferability without requiring extensive training data. As a result, finetuned models achieve superior performance in AMC.

$$y = f(Wx + b) \quad (2)$$

where  $x$  is the input feature vector,  $W$  is the learnable weight matrix,  $b$  is the bias vector, and  $f(\cdot)$  is a nonlinear activation function such as ReLU, defined as  $f(x) = \max(0, x)$ . This transformation consists of two key steps: a linear transformation ( $Wx + b$ ), followed by a nonlinear mapping through  $f(\cdot)$ , enabling the model to capture complex feature interactions. Adding dense layers enhances feature interaction by combining frequency–spatial dependencies, nonlinearity by using ReLU improves gradient flow, and generalization improves as deepened networks prevent feature loss at low SNRs.

$$W_{t+1} = W_t - \eta \nabla L(W_t) \quad (3)$$

where  $W_t$  represents the weight matrix at step  $t$ , while  $W_{t+1}$  is the updated weight matrix after a gradient descent step;  $\eta$  (learning rate) controls step size—smaller values ensure stability, whereas larger values speed up convergence but may cause instability;  $\nabla L$  is the gradient of the loss function  $L$  with respect to  $W_t$ , indicating the direction and magnitude of weight adjustments needed to minimize the loss.

$$X = \frac{x - \mu_B}{\sigma_B} \quad (4)$$

where  $x$  represents the input activation (output of the previous layer),  $\mu_B$  is the mean of activations within a batch,  $\sigma_B$  is the standard deviation of activations within a batch, and  $X$  is the normalized output after batch normalization.

$$\tilde{y} = M \circledast y, M \sim \text{Bernoulli}(p) \quad (5)$$

where  $y$  represents the original activation output before dropout,  $M$  is a binary mask sampled from a Bernoulli distribution with probability  $p$  (dropout rate),  $\tilde{y}$  is the final output after applying dropout, and  $\circledast$  denotes element-wise multiplication.  $M \sim \text{Bernoulli}(p)$  means  $M$  ( $\sim$  means is sampled) is sampled from a Bernoulli distribution with probability  $p$ .

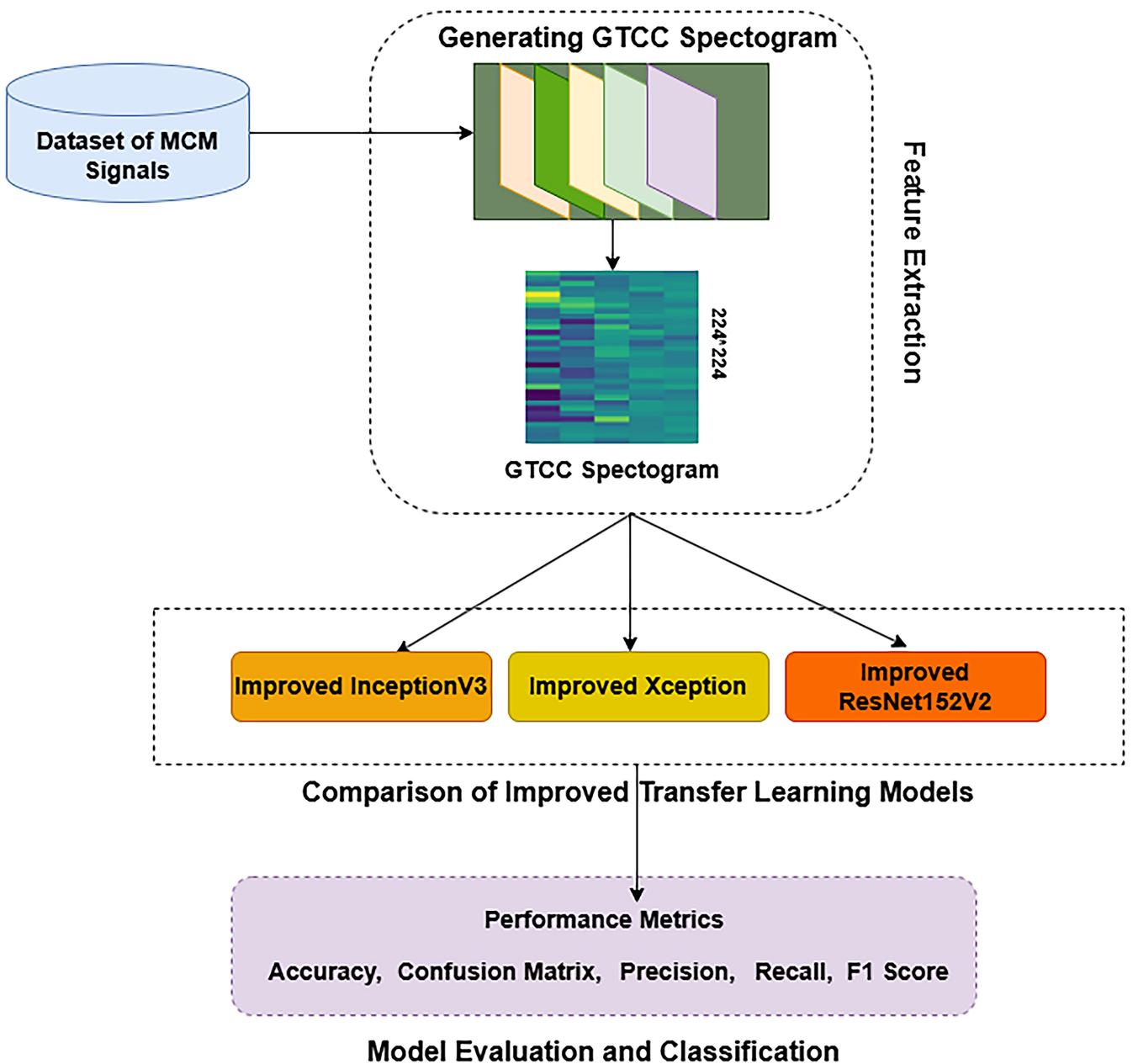
## 4 | Proposed Architecture

The proposed AMC architecture comprises two stages, as shown in Figure 2. In the first stage, GTCC spectrograms extract key features from MCM signals, providing a visual representation of their characteristics. In the second stage, the Improved-InceptionV3, Xception, and ResNet152V2 models, fine-tuned with these techniques, are used for classification, outperforming traditional approaches and making them well suited for next-generation wireless communication systems.

### 4.1 | Spectrogram Generation

The GTCC spectrogram is generated using a series of computational steps, as outlined in Function 1. This involves preprocessing the signal, applying the gammatone filter bank, computing energies, and performing DCT transformation to obtain the coefficients. Figure 3 provides a visual representation of the signal's spectral content over time, facilitating nuanced signal interpretation and feature extraction. Function 1 outlines a process for generating and saving GTCC spectrogram images from a CSV file containing MCM signal data. Initially, the signal data is loaded from a CSV file using “numpy.genfromtxt,” which reads the data while skipping the header. The data is then reshaped into a specified number of signals and subframes. For each signal, normalization is performed by dividing the signal by its maximum absolute value to ensure that the values are within a range suitable for further processing. A gammatone spectrogram (“gtgram”) is computed using parameters such as sample rate, window time, hop time, the number of filters, and the minimum frequency (“f\_min”). The logarithm of the gammatone spectrogram (“log\_gtgram”) is then taken to enhance the dynamic range. The DCT is applied to the log spectrogram to derive the GTCC features, focusing on a specific range of coefficients determined by “num\_ceps”. The resulting GTCC spectrogram is saved as a PNG image using “matplotlib.pyplot” with specified formatting options to ensure a tight layout and transparency.

The parameters used in the above GTCC spectrogram generation are shown in Table 2. These parameters define the setup for processing a signal using a gammatone filter bank, which is used in GTCC spectrogram generation. The sample rate is 44,100 Hz, meaning 44,100 samples are recorded per second, with a frequency range from 50 to 22,050 Hz (half the sample rate). Signal analysis is done in a 25-ms window, with a 10-ms overlap to prevent data loss between frames. The filter bank employs 32 filters, each targeting a specific frequency range, enabling detailed frequency decomposition for robust signal analysis.



**FIGURE 2 |** Proposed architecture.

This setup ensures smooth transitions between segments and comprehensive frequency extraction. In the context of MCM signal classification, fixing these parameters helps maintain a uniform representation of modulation characteristics in the GTCC spectrograms. This stability is critical for enabling deep learning models to learn discriminative patterns across modulation schemes without being influenced by variability in feature extraction.

## 4.2 | Classification

The proposed study's back-end implementation involves training and assessing MCM signal spectrograms using transfer learning models. The classification phase of the proposed architecture involves improved transfer learning models

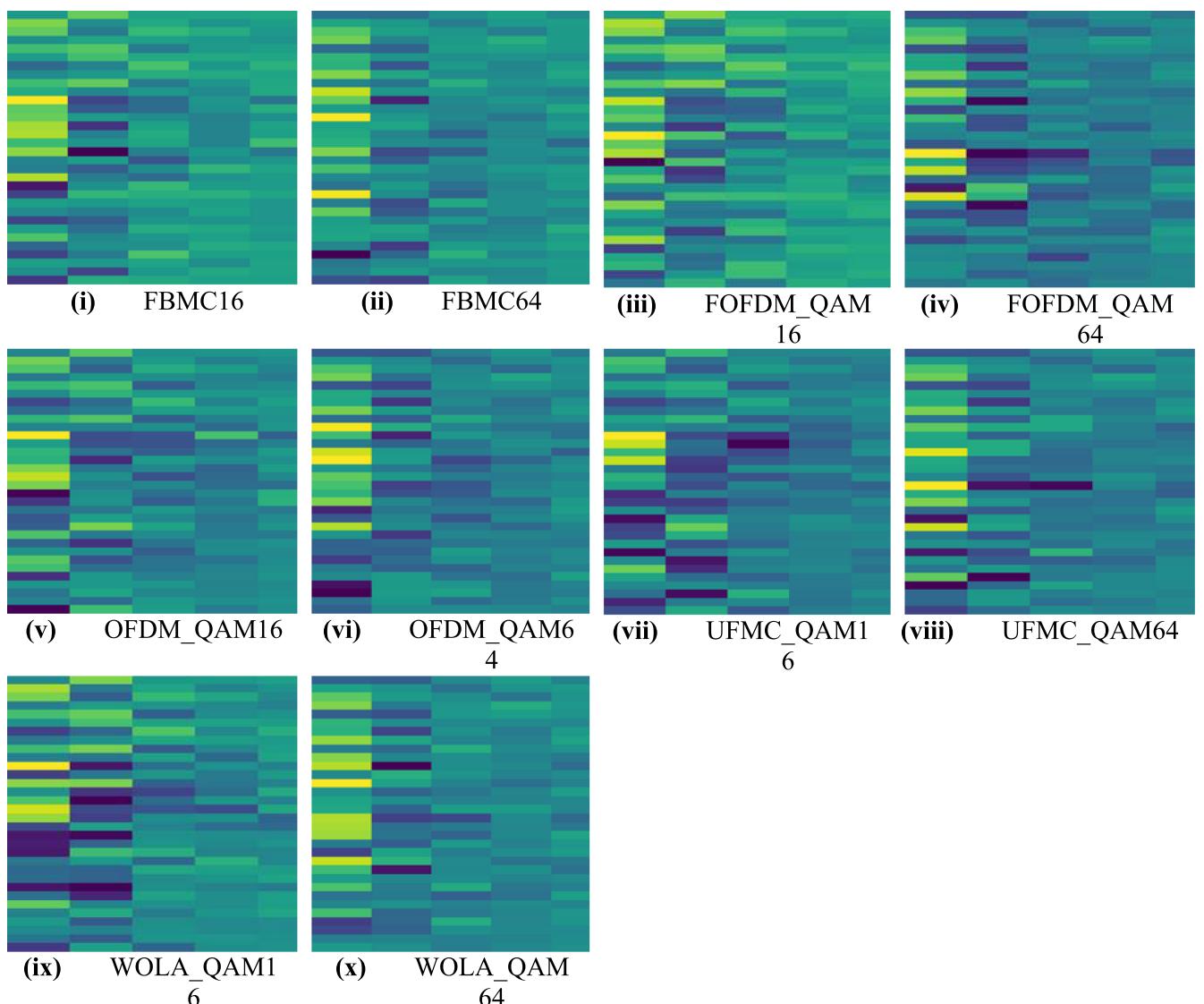
such as InceptionV3, Xception, and ResNet152V2 to classify GTCC spectrograms. These models are initialized with pre-trained weights and are designed to handle input images of shape (224, 224, and 3), corresponding to the dimensions of the GTCC spectrograms, leveraging TensorFlow's Keras API for implementation. Function 2 represents the algorithm for the Improved-InceptionV3 model, which is also used for two other improved models, differing only in the choice of the pre-trained model.

### 4.2.1 | Classification Using Improved-InceptionV3

Figure 4 illustrates the architecture of the Improved-InceptionV3 model, which employs transfer learning using a pretrained InceptionV3 with `include_top=False`. A custom classifier is

**FUNCTION 1** | GTCC spectrogram generation.**Input:** MCM Signals

1. *Input = "path of .csv file of MCM signal"*
2. *Output\_directory = "path to save generated GTCC spectrogram. PNG Image"*
3. *data = np.genfromtxt(Input\_filepath, delimiter = "", skip\_header = 1)*
4. *replicas = data.reshape((no. of signals, num\_subframes))*
5. *for i, row in replicas:*
6. *signal = replicas[i]*
7. *signal = signal / np.max(np.abs(signal))*
8. *gtgram\_matrix = gtgram(signal, sample\_rate, window\_time, hop\_time, num\_filters, f\_min)*
9. *log\_gtgram = np.log1p(gtgram\_matrix)*
10. *gtcc = dct(log\_gtgram, type = 2, axis = 1, norm = 'ortho')[ :, 1:(num\_ceps + 1) ]*
11. *gtcc\_features = compute\_gtcc(log\_gtgram, num\_ceps)*
12. *plt.savefig(resize, bbox\_inches = 'tight', pad\_inches = 0.0, transparent = True)*
13. *plt.close()*

**Output:** GTCC Spetrograms of MCM Signals**FIGURE 3** | Generated GTCC spectrograms of (i) FBMC16, (ii) FBMC64, (iii) FOFDM\_QAM16, (iv) FOFDM\_QAM64, (v) OFDM\_QAM16, (vi) OFDM\_QAM64, (vii) UFMC\_QAM16, (viii) UFMC\_QAM64, (ix) WOLA\_QAM16, and (x) WOLA\_QAM64.

appended, consisting of a GlobalAveragePooling2D layer, a dense layer with 1024 ReLU units, and a softmax output layer for multiclass classification. During initial training, the convolutional base is frozen to retain pretrained features, focusing training on the new classifier layers.

This two-stage transfer learning framework as described in Function 2 leveraging ImageNet pretrained hierarchical

**TABLE 2** | Parameters used for gammatone filter bank.

Variable name	Value
sample_rate	44,100
f_min (minimum frequency)	50
f_max (maximum frequency)	sample_rate/2
window_time (in seconds)	0.025
hop_time (in seconds)	0.010
num_filters (used in gammatone filter bank)	32

features followed by GTCC spectrogram-specific fine-tuning is pivotal for acoustic classification. By transferring cross-domain spatial representations, then optimizing them for GTCC's time-frequency nuances, the model captures discriminative auditory patterns while mitigating overfitting.

#### 4.2.2 | Classification Using Improved-Xception

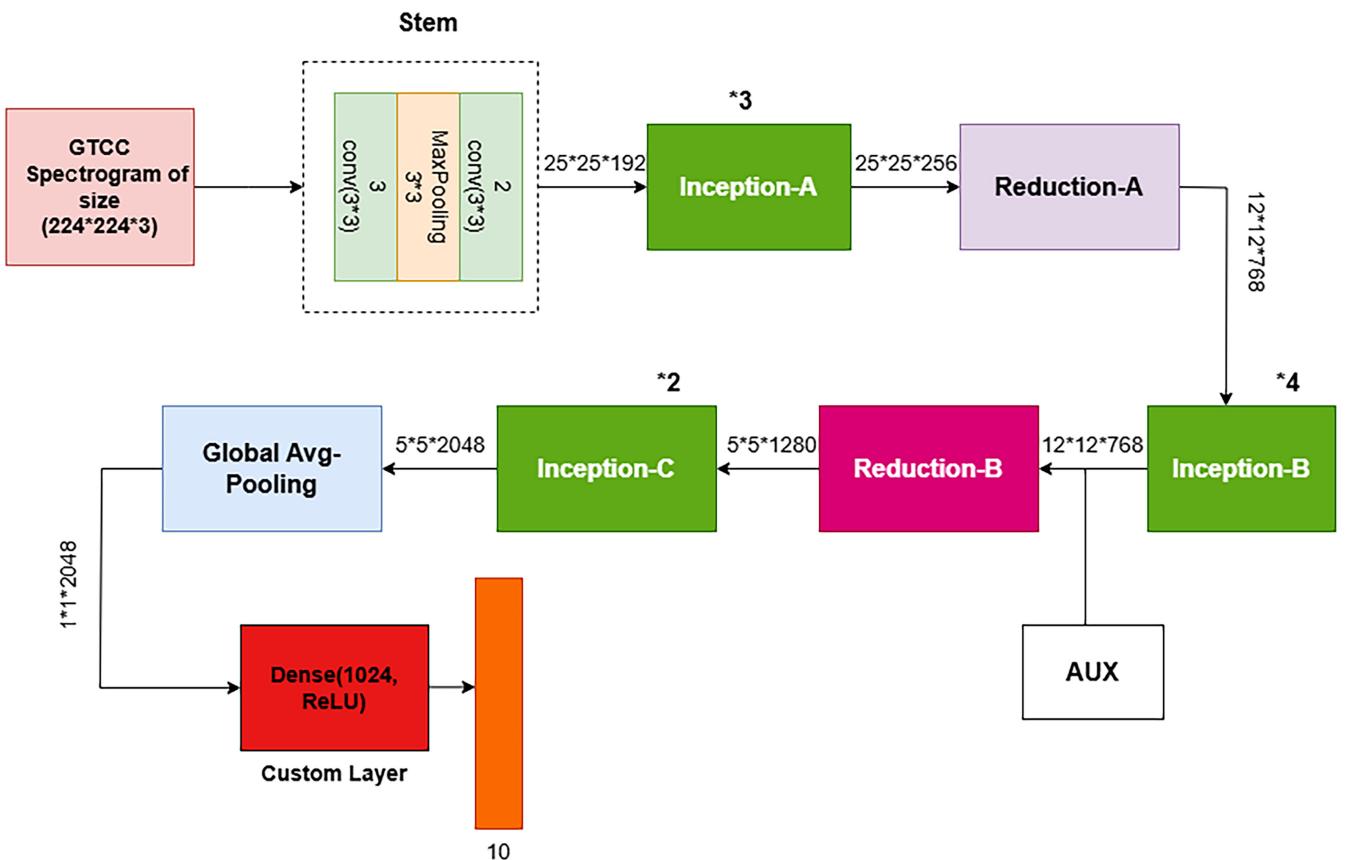
The architecture utilizes a pretrained Xception backbone to preserve hierarchical time-frequency abstractions ( $7 \times 7 \times 2048$  feature maps) from input spectrograms. A GAP layer collapses spatial dimensions to  $1 \times 2048$ , followed by a Dense(1024-ReLU) layer to distill modulation-specific embeddings, and a softmax layer for multicarrier class probabilities, as given in Function 2. By retaining Xception's depthwise separable convolutions, the model efficiently captures cross-domain spatial hierarchies critical for discriminating overlapping subcarriers and time-frequency sparsity patterns. The custom classifier then fine-tunes high-level representations to adapt to subtle multicarrier distortions while preserving pretrained feature robustness. Figure 5 represents the architecture of the Improved-Xception model.

#### FUNCTION 2 | Building of improved transfer learning model.

```

Input: training data = train, validation data = valid
1. #Initialize Parameters
2. num_classes = 10
3. learning_rate = 0.0001
4. #Load Pretrained Model (edit name for other 2 models)
5. base_model = InceptionV3(weights='imagenet', include_top=False, input_shape = 224 * 224)
6. #Add custom Layer
7. x = base_model.output
8. x = GlobalAveragePooling2D()(x)
9. x = Dense(1024, activation = 'relu')(x)
10. predictions = Dense(num_classes, activation = 'softmax')(x)
11. #Create Final Model
12. model = Model(inputs = base_model.input, outputs = predictions)
13. #Freeze base model layer
14. for each layer in base_model:
15. set layer.trainable = False
16. #Compile the model
17. model.compile(optimizer=Adam(learning_rate=learning_rate),
   loss=sparse_categorical_crossentropy, metrics=['accuracy'])
18. #Train the Model with Frozen Layers
19. model.fit(train, validation_data = valid, epochs = 10, batch_size = 32)
20. #Fine – Tuning by Unfreezing Base Model Layers
21. for each layer in base_model:
22. set layer.trainable = True
23. #Re – Compile the Model (by decreased learning rate)
24. model.compile(optimizer=Adam(learning_rate=learning_rate/10),
   loss=sparse_categorical_crossentropy, metrics=['accuracy'])
25. #Train the Improved model
26. model.fit(train, validation_data = valid, epochs = 10, batch_size = 32)
Output: Trained Improved Model

```



**FIGURE 4** | Proposed architecture of Improved-InceptionV3.

#### 4.2.3 | Classification Using Improved-ResNet152V2

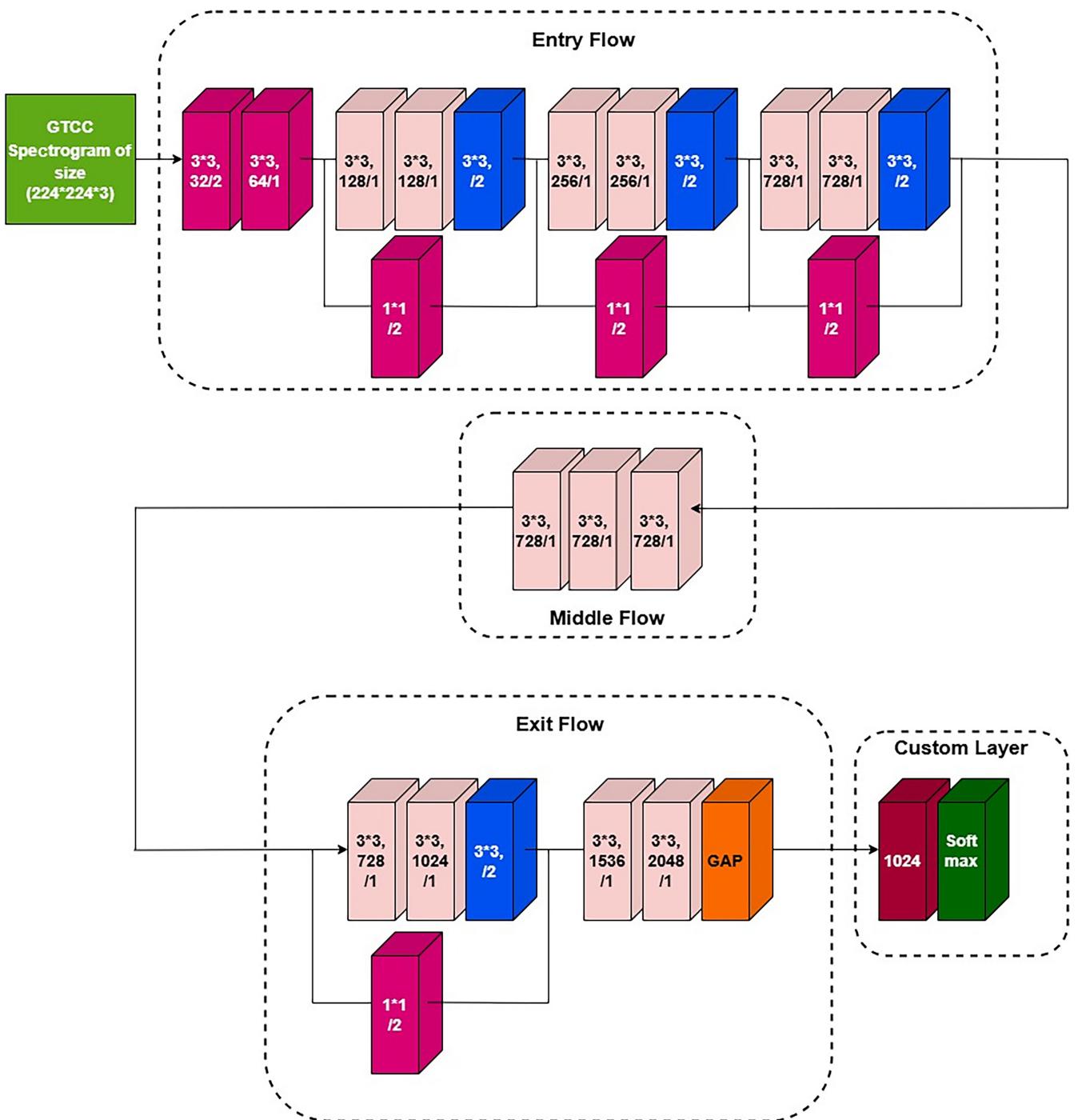
This model is also constructed using the same strategies outlined in the InceptionV3 architecture, such as utilizing pretrained weights and efficient feature extraction. However, instead of leveraging the InceptionV3 pretrained model, the ResNet152V2 architecture is used.

The Improved-ResNet152V2 is uniquely suited for classifying 5G MCM signals using GTCC spectrograms due to its deep hierarchical architecture and adaptive feature refinement. The model's four-stage residual blocks, with progressive spatial downsampling ( $56 \times 56 \rightarrow 7 \times 7$ ) and channel depth expansion ( $64 \rightarrow 2048$ ), capture fine-grained time-frequency patterns inherent to 5G waveforms, such as cyclic prefix variations, spectral leakage, and filterbank-induced distortions. Preactivation residual connections stabilize gradient propagation across its 152 layers, enabling robust learning of long-range spectral dependencies critical for distinguishing densely overlapping subcarriers in 5G environments. The two-phase training, as given in Function 2, involves first freezing ImageNet-pretrained low-/mid-level filters to preserve spatial hierarchies, then fine-tuning high-level layers to optimize adaptation to GTCC-specific artifacts while avoiding catastrophic forgetting. The architecture of this model is described in Figure 6.

## 5 | Experimental Setup and Results

This section outlines the tests that were carried out to assess the performance of our MCM signal categorization approach. Using a dataset of 10 MCM communication signals, we evaluated the methodology. The back-end model was implemented using Google Colab, utilizing the TensorFlow and Keras frameworks, and was built in Python 3.7. One of the computing resources used was a 15GB GDDR6 NVIDIA Tesla T4 GPU. The proposed models utilized transfer learning with pretrained ImageNet weights. During the initial training phase, the base model layers were frozen, and only the custom classification head was trained using the Adam optimizer with a learning rate of 0.0001, a batch size of 32, and for 10 epochs. In the second phase, the entire model was unfrozen for fine-tuning with a reduced learning rate divided by 10, and training continued for an additional 20 epochs, as given in Table 3. Evaluation criteria like accuracy (percentage), recall (percentage), precision (percentage), and *F1* score (percentage), as given by equations, are used to gauge how well the suggested model performs.

Accuracy measures overall prediction correctness, precision evaluates the reliability of positive predictions, recall quantifies the detection of true positives, and *F1* score balances precision and recall via their harmonic mean. Those equations are given below:



**FIGURE 5** | Proposed architecture of Improved-Xception.

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \quad (6)$$

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \quad (7)$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{TN}) \quad (8)$$

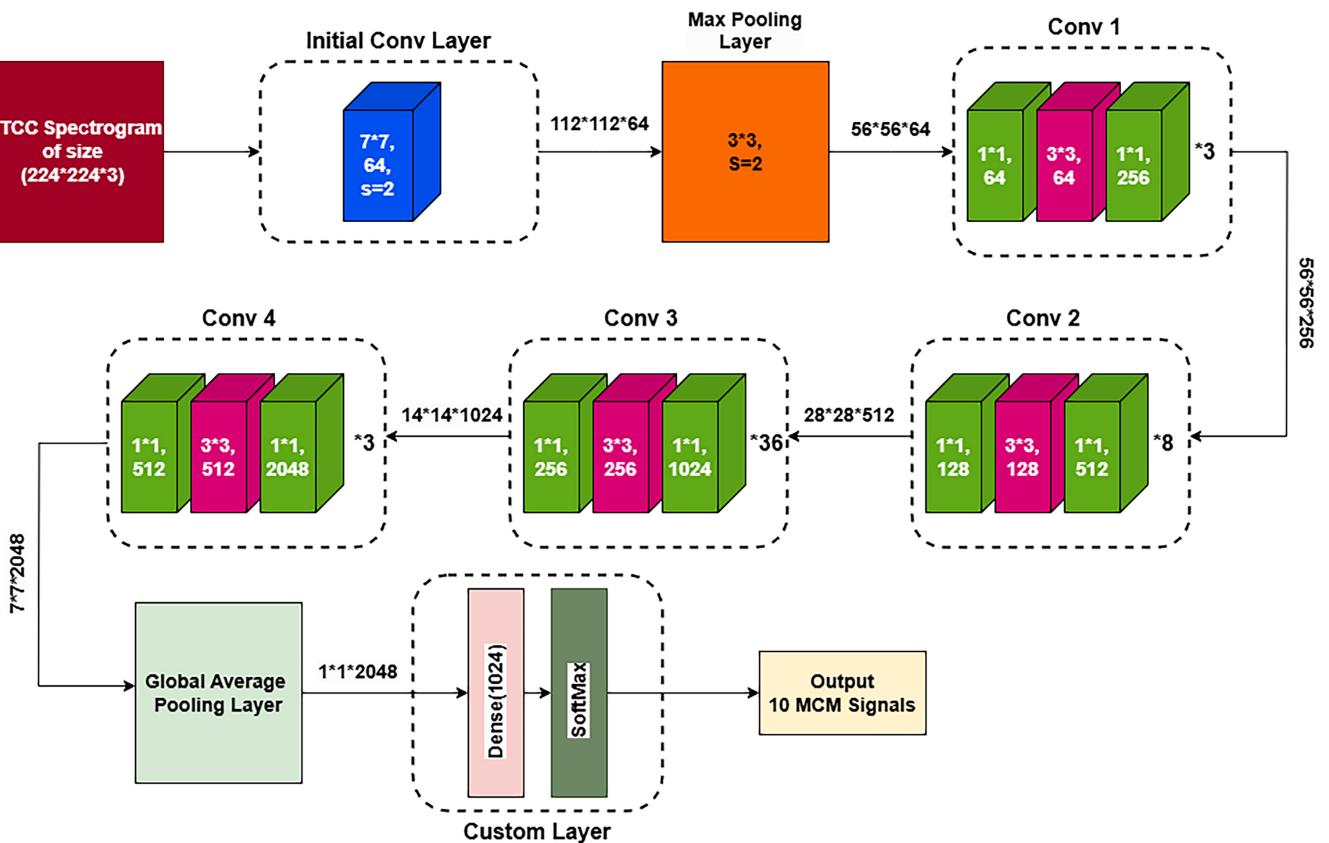
$$F1 - \text{Score} = (2 * \text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (9)$$

where TP is true positive, the number of correctly predicted positive samples; FP is false positive, the number of negative samples

incorrectly predicted as positive; and FN is false negative, the number of positive samples incorrectly predicted as negative.

## 5.1 | Dataset

A simulated dataset is utilized in this study, specifically crafted to address the challenges associated with spectrum sensing and MCM signal classification. The dataset emulates real-world conditions and signal variations typical in wireless



**FIGURE 6** | Proposed architecture of Improved-ResNet152V2.

**TABLE 3** | Implementation parameters.

Parameter	Value
Hardware platform for spectrogram generation	AMD Ryzen 75,700 U @ 1.80 GHz, 16 GB RAM
Backend platform	Google Colab with GPU (Tesla T4) for classification
Transfer learning	Pretrained ImageNet weights
Optimizer	Adam
Loss function	Sparse categorical cross-entropy
Initial learning rate	0.0001
Batch size	32
Epochs (Phase 1)	10
Model state (Phase 1)	Base model frozen
Learning rate (Phase 2)	0.00001
Epochs (Phase 2)	20
Model state (Phase 2)	Entire model unfrozen

communication environments, providing a robust basis for evaluating and developing advanced signal processing and classification algorithms.

The dataset contains 10 distinct MCM types, with signal samples generated across a wide range of SNRs from  $-20$  to  $+30$  dB, reflecting both noisy and clean communication scenarios. Each sample is explicitly labeled with its corresponding modulation type and SNR value, facilitating targeted analysis and performance tracking across varying conditions. Table 4 provides the dataset specifications.

To further replicate real-world noise situations, several degrees of white Gaussian noise are added. The dataset [13] includes simulated radio signals generated using the Vienna 5G Link Level Simulator [27] for training machine learning models in MCM detection and spectrum sensing. Random bit sequences are transmitted through a simulated free-space radio channel. The channel model incorporates frequency-selective fading with the Vehicular A [28] power delay profile and Jake's Doppler model to account for time selectivity. White Gaussian noise is added at varying levels to simulate different SNRs.

In this work, 50 GTCC spectrogram images were generated for each of the 10 modulation types, resulting in a total of 500 spectrograms per SNR level. These spectrograms serve as rich input representations for the classification task and were systematically organized into three separate directories: training, validation, and testing for each SNR value. The training and validation datasets contain evenly distributed samples across all classes, while the testing set is created using an 80:20 stratified split. Prior to training, labels were encoded using standard encoding techniques to ensure compatibility with the classification models.

Sections 5.2.1, 5.2.2, and 5.2.3 present the evaluation of different models to classify GTCC spectrograms of MCM signals, including Improved-InceptionV3, Improved-Xception, and Improved-ResNet152V2, providing a performance analysis of their accuracy across a range of SNR values from -16 to 16 dB.

## 5.2 | Computational Resource Requirements and Runtime Analysis

The comparative analysis of the models highlights that *Improved-ResNet152V2* achieves the fastest inference time of 0.1048 s per image, outperforming both *Improved-InceptionV3* and *Improved-Xception*. Although it has a larger model size (693.69 MB) and total parameter footprint (246.65 MB), it maintains the same trainable parameter size (8.04 MB) as the others, indicating a controlled training overhead. The increased non-trainable parameters contribute to its depth and feature extraction capability, which is crucial for capturing fine-grained details in GTCC spectrograms. Table 5 shows its balance of speed, high representational capacity, and superior performance. *Improved-ResNet152V2* proves to be a reliable and optimal choice for modulation classification using GTCC features, especially in scenarios where accuracy and robust feature learning are prioritized.

### 5.2.1 | Evaluation of GTCC Spectrograms With Improved-InceptionV3 Model

Table 6 presents the classification report for the fine-tuned InceptionV3 model across a SNR range from -16 to 16 dB. The

model demonstrates a classification accuracy of 88% at -16 dB, which remains consistent at 88% at -12 dB. As the SNR increases, the classification accuracy improves, reaching 93% at -8 dB, 96% at -4 dB, and 97% at 0 dB. The model maintains an accuracy of 97% at 4, 8, and 12 dB and further improves to 98% at 16 dB. These results indicate that the model achieves higher classification accuracy at higher SNR levels, with the peak performance observed at 16 dB.

### 5.2.2 | Evaluation of GTCC Spectrograms With Improved-Xception Model

Table 7 shows the classification performance of the Improved-Xception model across SNR range from -16 to 16 dB. The model's accuracy improves from 86% at -16 dB to 94% at -12 dB, reaching 95% at -8 dB and 97% at -4 dB, remaining consistent at 97% from 0 to 4 dB. Accuracy slightly decreases to 96% at 8 dB, then returns to 97% at 12 dB and stays at 97% at 16 dB. These results indicate that the Improved-Xception model achieves higher classification accuracy at increased SNR levels, with stable high performance at SNRs of 4 dB and above.

### 5.2.3 | Evaluation of GTCC Spectrograms With Improved-ResNet152V2 Model

Table 8 shows the classification performance of the Improved-ResNet152V2 model across an SNR range from -16 to 16 dB. The model maintains an accuracy of 96% from -16 to -12 dB, improves to 98% at -8 dB, and reaches 99% from -4 to 0 dB. Accuracy further increases to 100% from 4 to 16 dB. These

**TABLE 4** | 5G waveform dataset [13].

No. of modulation signals	10
Signal's name	FBMC_QAM16, FBMC_QAM64, OFDM_QAM16, FOFDM_QAM64, OFDM_QAM16, OFDM_QAM64, WOLA_QAM16, WOLA_QAM64, UFMC_QAM16, and UFMC_QAM64
No. of signals (in each SNR file)	1000
No. of subframes (columns for each signal)	3627 (for FOFDM_QAM16 and QAM64 and WOLA_QAM16 and WOLA_QAM64) 3626 (for remaining all signals)
SNR range	-16:4:16

**TABLE 5** | Comparison of inference time, model size, and parameters for improved models.

Models	Inference time per images (in second)	Model's size (in MB)	Trainable param (in MB)	Nontrainable param (in MB)	Total param (in MB)
Improved-InceptionV3	0.1286	274.49	8.04	83.17	107.30
Improved-Xception	0.1077	263.03	8.04	79.58	103.71
Improved-ResNet152V2	0.1048	693.69	8.04	222.52	246.65

**TABLE 6** | Classification report of Improved-InceptionV3.

SNR	Metric	WOLA		UFMC		OFDM		FOFDM		FBMC	
		QAM16	QAM64								
-16	Precision	1.00	0.82	0.82	1.00	0.75	1.00	0.90	0.83	0.88	0.91
	Recall	0.80	0.90	0.90	0.70	0.90	1.00	0.90	1.00	0.70	1.00
	F1 score	0.82	0.86	0.86	0.82	0.82	1.00	0.90	0.91	0.78	0.95
-12	Precision	0.90	0.67	1.00	0.90	1.00	1.00	1.00	1.00	0.75	0.90
	Recall	0.90	0.80	0.90	0.90	0.90	0.90	0.80	0.80	0.90	0.90
	F1 score	0.90	0.73	0.95	0.90	0.95	0.95	0.89	0.89	0.82	0.90
-8	Precision	1.00	1.00	0.82	1.00	1.00	1.00	0.77	1.00	0.90	0.91
	Recall	0.90	1.00	0.90	1.00	0.90	0.80	1.00	0.90	0.90	1.00
	F1 score	0.95	1.00	0.86	1.00	0.95	0.89	0.87	0.95	0.90	0.95
-4	Precision	1.00	1.00	0.82	1.00	1.00	1.00	0.77	1.00	0.90	0.91
	Recall	0.90	1.00	0.90	1.00	0.90	0.80	1.00	0.90	0.90	1.00
	F1 score	0.95	1.00	0.86	1.00	0.95	0.89	0.87	0.95	0.90	0.95
0	Precision	1.00	0.91	0.91	1.00	1.00	1.00	0.91	1.00	1.00	1.00
	Recall	1.00	1.00	1.00	1.00	0.80	0.90	1.00	1.00	0.90	1.00
	F1 score	1.00	0.95	0.95	1.00	0.89	0.95	0.95	1.00	0.95	1.00
4	Precision	1.00	0.91	0.91	1.00	1.00	1.00	1.00	0.91	1.00	1.00
	Recall	0.90	1.00	1.00	1.00	0.80	1.00	1.00	1.00	1.00	1.00
	F1 score	0.95	0.95	0.95	1.00	0.89	1.00	1.00	0.95	1.00	1.00
8	Precision	0.91	1.00	1.00	1.00	0.91	1.00	1.00	1.00	0.91	1.00
	Recall	1.00	1.00	0.90	1.00	1.00	1.00	0.90	1.00	1.00	0.90
	F1 score	0.95	1.00	0.95	1.00	0.95	1.00	0.95	1.00	0.95	0.95
12	Precision	1.00	0.90	0.91	1.00	1.00	0.91	1.00	1.00	1.00	1.00
	Recall	0.90	0.90	1.00	1.00	1.00	1.00	1.00	1.00	0.90	1.00
	F1 score	0.95	0.90	0.95	1.00	1.00	0.95	1.00	1.00	0.95	1.00
16	Precision	0.91	1.00	1.00	0.91	1.00	1.00	1.00	1.00	1.00	1.00
	Recall	1.00	0.90	1.00	1.00	1.00	1.00	0.90	1.00	1.00	1.00
	F1 score	0.95	0.95	1.00	0.95	1.00	1.00	0.95	1.00	1.00	1.00

results demonstrate that the Improved-ResNet152V2 model achieves highest classification accuracy as compare to other two models.

The evaluation of three models demonstrates that Improved-ResNet152V2 outperforms the others in the classification of GTCC spectrograms of MCM signals, even at negative SNR values. Figure 7 provides an additional analysis using a confusion matrix, further highlighting the superior performance of Improved-ResNet152V2 compared to the other models. Improved-ResNet152V2 effectively handles diverse signal conditions due to its deep residual structure, which supports learning robust and discriminative features across varying signal types and SNR levels. The use of pretrained ImageNet weights enables the model to benefit from generalized feature

representations, while domain-specific fine-tuning on GTCC spectrograms allows adaptation to the unique characteristics of MCM signals. The inclusion of a GAP layer reduces model complexity and overfitting risk by minimizing the number of parameters, which further contributes to stable performance across fluctuating noise environments and modulation formats.

### 5.3 | Robustness Analysis via Noisy Inputs and Grad-CAM Visualization

To evaluate the robustness of the Improved-ResNet152V2 model, colored Gaussian noise was synthetically added to input spectrogram images. This noise simulates real-world

**TABLE 7** | Classification report of Improved-Xception model.

SNR	Metric	WOLA		UFMC		OFDM		FOFDM		FBMC	
		QAM16	QAM64								
-16	Precision	1.00	1.00	0.75	1.00	0.62	0.70	0.88	0.70	0.89	0.91
	Recall	0.80	0.90	0.90	1.00	0.80	1.00	0.70	1.00	0.80	1.00
	F1 score	0.89	0.95	0.82	1.00	0.70	0.87	0.78	0.87	0.84	0.95
-12	Precision	0.90	1.00	1.00	1.00	0.77	1.00	0.90	1.00	0.88	1.00
	Recall	0.90	1.00	1.00	1.00	1.00	0.90	0.90	1.00	0.70	1.00
	F1 score	0.90	1.00	1.00	1.00	0.87	0.95	0.90	1.00	0.78	1.00
-8	Precision	1.00	1.00	0.91	0.91	1.00	1.00	1.00	1.00	1.00	0.77
	Recall	1.00	1.00	1.00	1.00	0.90	0.90	0.90	0.80	1.00	1.00
	F1 score	1.00	1.00	0.95	0.95	0.95	0.95	0.95	0.89	1.00	0.87
-4	Precision	1.00	1.00	1.00	1.00	0.91	0.83	1.00	1.00	1.00	1.00
	Recall	1.00	1.00	1.00	0.90	1.00	1.00	1.00	0.90	1.00	0.90
	F1 score	1.00	1.00	1.00	0.95	0.95	0.91	1.00	0.95	1.00	0.95
0	Precision	1.00	1.00	0.91	1.00	1.00	1.00	0.83	1.00	1.00	1.00
	Recall	0.70	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1 score	0.82	1.00	0.95	1.00	1.00	1.00	0.91	1.00	1.00	1.00
4	Precision	1.00	1.00	0.91	1.00	0.90	0.91	1.00	1.00	1.00	1.00
	Recall	0.90	1.00	1.00	1.00	0.90	1.00	0.90	1.00	1.00	1.00
	F1 score	0.95	1.00	0.95	1.00	0.90	0.95	0.95	1.00	1.00	1.00
8	Precision	1.00	0.91	0.91	1.00	1.00	1.00	0.90	1.00	0.89	1.00
	Recall	1.00	1.00	1.00	1.00	1.00	1.00	0.90	0.90	0.80	1.00
	F1 score	1.00	0.95	0.95	1.00	1.00	1.00	0.90	0.95	0.84	1.00
12	Precision	1.00	1.00	1.00	0.83	0.91	1.00	1.00	1.00	1.00	1.00
	Recall	1.00	0.90	1.00	1.00	1.00	1.00	0.80	1.00	1.00	1.00
	F1 score	1.00	0.95	1.00	0.91	0.95	1.00	0.89	1.00	1.00	1.00
16	Precision	1.00	1.00	1.00	1.00	0.91	1.00	1.00	0.90	1.00	0.90
	Recall	1.00	1.00	1.00	0.90	1.00	1.00	1.00	0.90	1.00	0.90
	F1 score	1.00	1.00	1.00	0.95	0.95	1.00	1.00	0.90	1.00	0.90

channel perturbations by introducing structured distortions across color channels. Despite these modifications, the Grad-CAM visualizations of noisy inputs show that the model consistently focuses on the same discriminative regions as in the clean inputs. This consistency highlights the model's ability to maintain feature localization and classification performance even under degraded signal conditions, demonstrating resilience to noise and confirming the stability of learned representations.

Figure 8 presents Grad-CAM visualizations for correctly classified spectrograms by the Improved-ResNet152V2 model. The original spectrograms and their corresponding Grad-CAMs

highlight the discriminative regions responsible for accurate classification. Even after the addition of colored noise, the model consistently focuses on the same key features, demonstrating its robustness and effective feature localization under noisy conditions.

#### 5.4 | Overhead Analysis

This section shows the computational overhead across the experimental setups of the compared works as given in Table 9. Overhead includes training and inference complexity, hardware requirements, and deployment practicality, all of which are

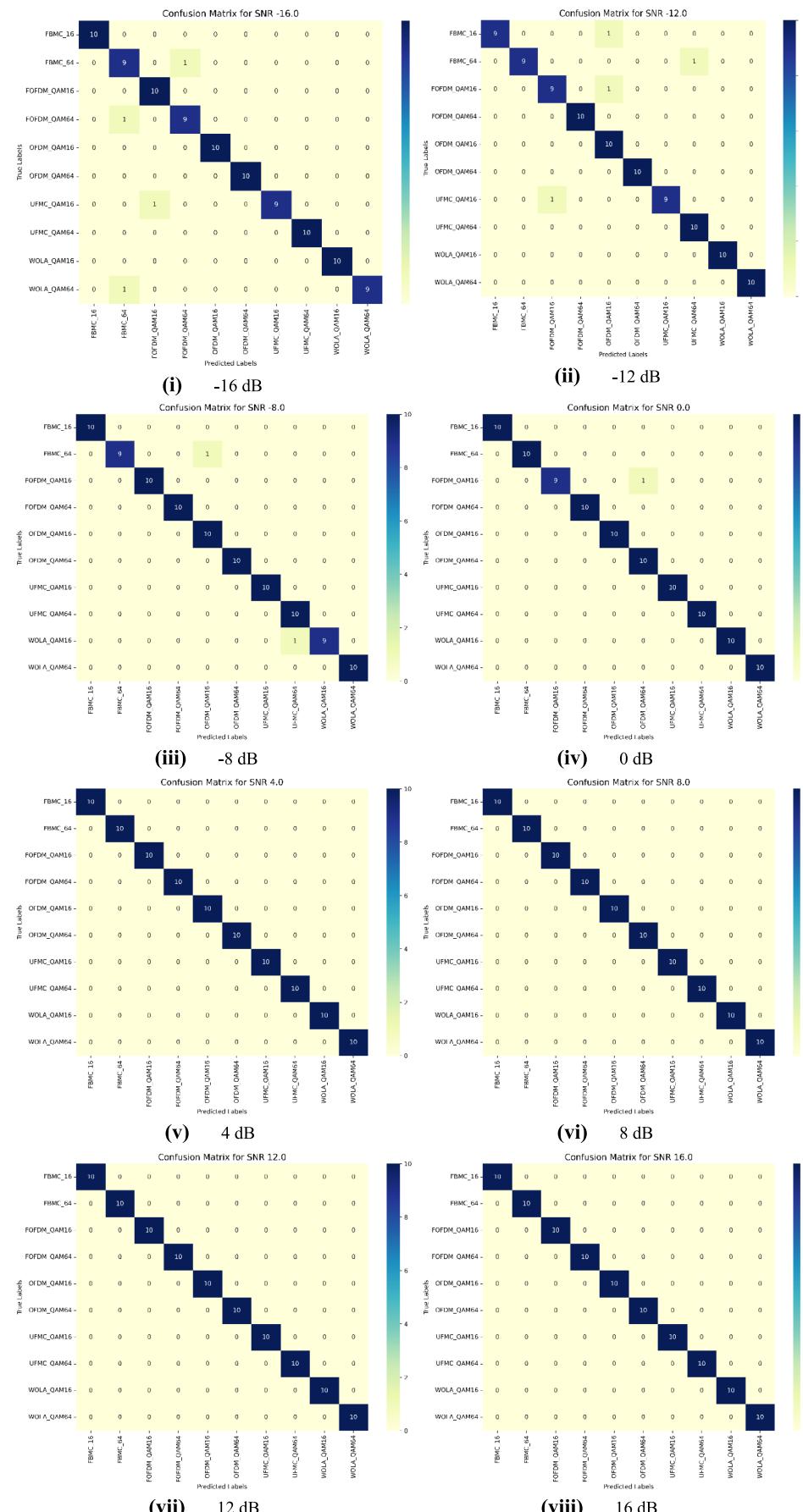
**TABLE 8** | Classification report of Improved-ResNet152V2 model.

SNR	Metric	WOLA		UFMC		OFDM		FOFDM		FBMC	
		QAM16	QAM64								
-16	Precision	1.00	1.00	1.00	1.00	1.00	1.00	0.91	0.90	1.00	0.82
	Recall	1.00	0.90	0.90	1.00	1.00	1.00	1.00	0.90	1.00	0.90
	F1 score	1.00	0.95	0.95	1.00	1.00	1.00	0.95	0.90	1.00	0.86
-12	Precision	1.00	1.00	1.00	0.91	0.83	1.00	0.90	1.00	1.00	1.00
	Recall	1.00	1.00	0.90	1.00	1.00	1.00	0.90	1.00	0.90	0.90
	F1 score	1.00	1.00	0.95	0.95	0.91	1.00	0.90	1.00	0.95	0.95
-8	Precision	1.00	1.00	1.00	0.91	0.91	1.00	1.00	1.00	1.00	1.00
	Recall	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.90
	F1 score	0.95	1.00	1.00	0.95	0.95	1.00	1.00	1.00	1.00	0.95
-4	Precision	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.91	1.00	1.00
	Recall	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1 score	0.95	1.00	1.00	1.00	1.00	1.00	1.00	0.95	1.00	1.00
0	Precision	1.00	1.00	1.00	1.00	1.00	0.91	1.00	1.00	1.00	1.00
	Recall	1.00	1.00	1.00	1.00	1.00	1.00	0.90	1.00	1.00	1.00
	F1 score	1.00	1.00	1.00	1.00	1.00	0.95	0.95	1.00	1.00	1.00
4	Precision	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Recall	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1 score	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
8	Precision	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Recall	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1 score	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
12	Precision	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Recall	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1 score	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
16	Precision	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Recall	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F1 score	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

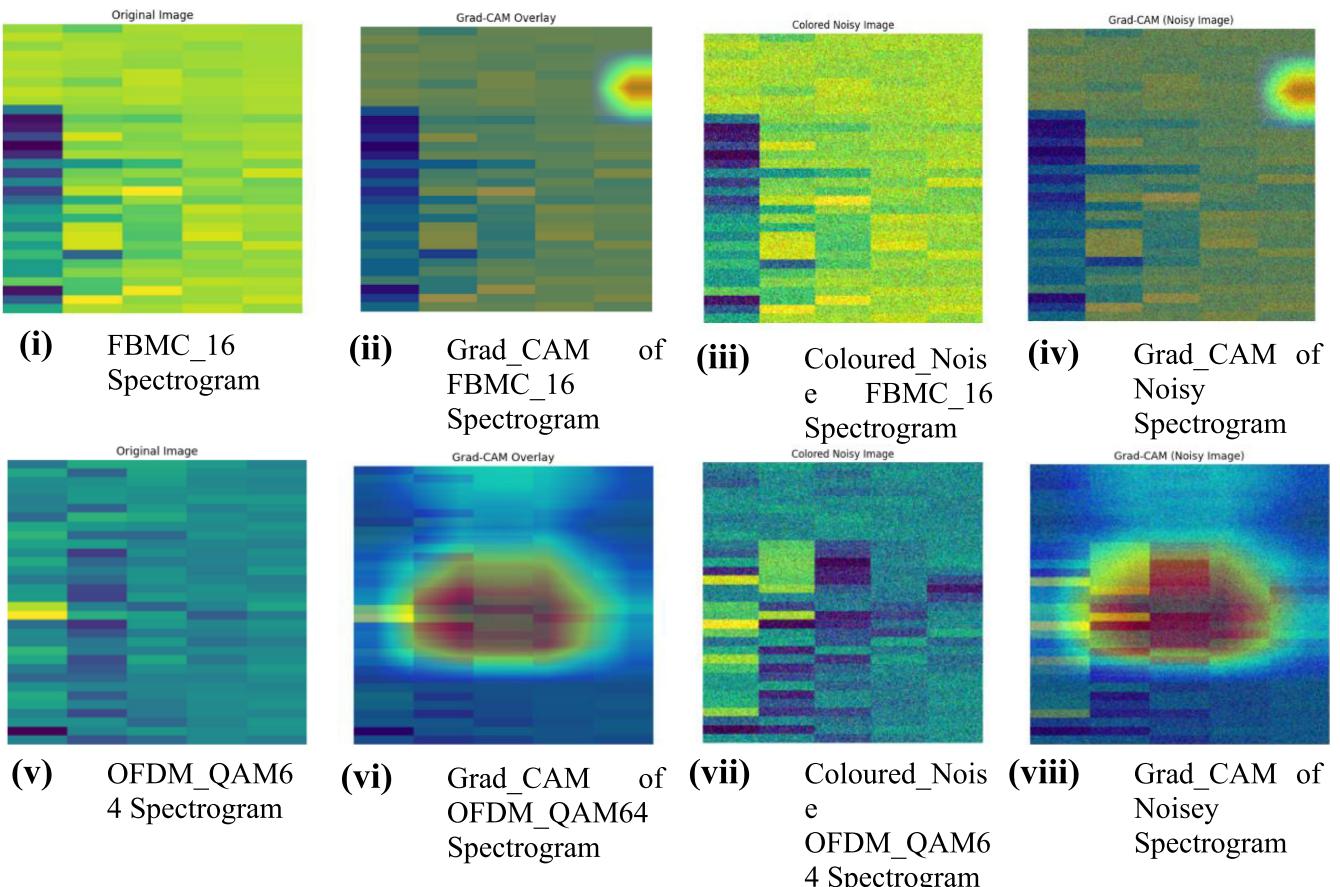
crucial for real-world applications such as cognitive radios and spectrum monitoring systems.

In comparison to existing approaches that rely on conventional spectrogram generation techniques like STFT, which involve higher preprocessing time and GPU-dependent training pipelines, the proposed method integrates a lightweight GTCC-based preprocessing strategy with optimized deep convolutional networks, where in computing complexity of GTCC  $N$  refers to number of samples per signal. While Duan et al. and Wu et al. shows training times of approximately 20–25 min per epoch with over 1 million parameters and complexities in the order of  $O(n \cdot k^2 \cdot f)$ , where  $n$  refers to input feature size,  $f$  refers to number of filters per layer, and  $k$  refers

to kernel size of convolution filters. The proposed back-end models, though analyzed for complexity in Table 5, demonstrate clear computational advantages over existing methods. For instance, GTCC combined with Improved-Xception achieves an inference time of 0.1777 s per image, compared to the higher overhead of STFT-based pipelines. Additionally, the theoretical complexity of the proposed pipeline remains efficient due to faster spectral representation and optimized model structures such as depthwise separable convolutions and residual blocks. Overall, the proposed framework outperforms existing methods by offering a better trade-off between computational efficiency and classification performance, making it more suitable for real-time and large-scale modulation recognition tasks.

**FIGURE 7 |** Legend on next page.

**FIGURE 7** | Confusion matrix of Improved-ResNet152V2 model at SNR values: (i) −16, (ii) −12, (iii) −8, (iv) 0, (v) 4, (vi) 8, (vii) 12, and (viii) 16 dB.



**FIGURE 8** | Grad-CAM of original spectrograms and noisy spectrograms of correctly predicted images on Improved-ResNet152V2 Model: (i) FBMC16 spectrogram, (ii) Grad-CAM FBMC16 spectrogram, (iii) Coloured\_Noise FBMC16 spectrogram, (iv) Grad-CAM noisy spectrogram, (v) OFDM\_QAM64 spectrogram, (vi) Grad-CAM of OFDM\_QAM64 spectrogram, (vii) Coloured\_Noise OFDM\_QAM64 spectrogram, and (viii) Grad-CAM noisy spectrogram.

## 6 | Discussion and Comparative Analysis

The comparative evaluation of the three models Improved-InceptionV3, Improved-Xception, and Improved-ResNet152V2 across varying SNR levels reveals distinct performance trends in the classification of GTCC spectrograms of MCM signals. As depicted in Table 6, the Improved-InceptionV3 model demonstrates a steady increase in classification accuracy from 88% at −16 and −12dB to 98% at 16dB, indicating a strong capability to handle higher SNR levels, with optimal performance observed at 16dB. Similarly, the Improved-Xception model, as shown in Table 7, starts with an accuracy of 86% at −16dB, improving significantly to 94% at −12dB and peaking at 97% from −4 to 0dB. The model maintains high performance at SNRs above 0dB, with a slight dip to 96% at 8dB before stabilizing again at 97% up to 16dB. This consistency demonstrates Improved-Xception's robust performance across a wide range of SNR levels, particularly from 4dB upwards.

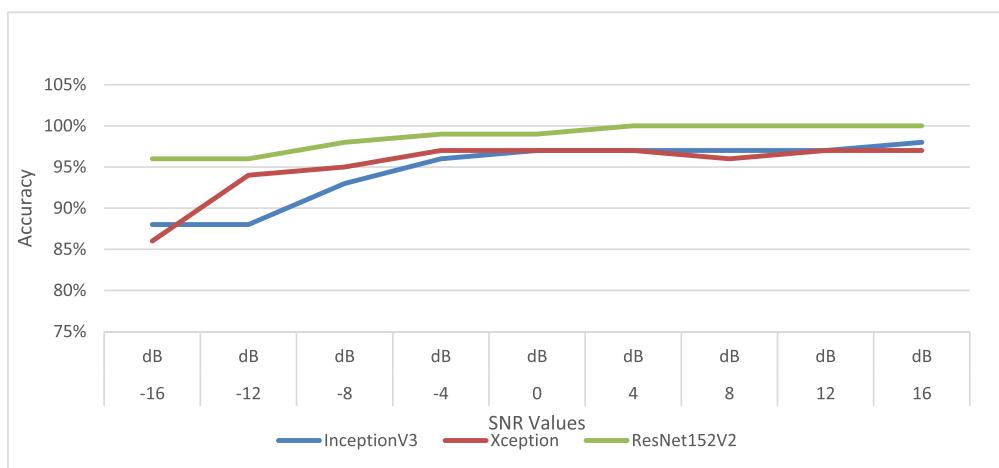
Table 8 shows that the Improved-ResNet152V2 model consistently outperforms Improved-InceptionV3 and Improved-Xception across all SNR levels to classify the GTCC spectrograms,

maintaining a high accuracy of 96% at −16 and −12dB, reaching 98% at −8dB, and achieving 99% from −4 to 0dB, with a perfect accuracy of 100% from 4 to 16dB. These results suggest that Improved-ResNet152V2 is more effective at sustaining high classification accuracy, even under challenging conditions with negative SNR values. Figure 9 and Table 10 further confirm that Improved-ResNet152V2 is the most reliable model for classifying GTCC spectrograms of MCM signals, particularly in low SNR scenarios, due to its superior performance across all SNR ranges. This underscores the advantage of deeper networks with identity mapping, such as Improved-ResNet152V2, in handling diverse signal conditions.

The existing and proposed works are not implemented under identical experimental conditions as given in Table 8. Although all approaches employ a front-end and deep learning-based back-end for modulation classification, they differ in datasets, SNR ranges, and feature extraction methods. Notably, only Tailor et al. [5] use the same 5G waveform dataset, making it the most relevant for comparison. While some prior works use STFT-based spectrograms, the proposed method introduces

**TABLE 9** | Overhead analysis with existing approaches and dataset.

Work	Dataset	Platform	Model	Computational cost	Computing complexity
Duan et al. [9]	227,940 STFT spectrograms (16 modulation types)	16 GB RAM, 2×NVIDIA Titan V100 GPUs, TensorFlow 1.12, Keras 2.2	CNN	~20 min/epoch (GPU), ~1 M parameters	$O(n \cdot k^2 \cdot f)$
Xie et al. [2]	Simulated signals	Offline platform (unspecified)	4-layer DNN	< 10 min total, ~800 parameters	$O(n \cdot h)$ (linear layers, few neurons)
Wu et al. [4]	16 modulation types	Google Colab: NVIDIA Tesla T4 (15 GB GPU)	CNN	~25 min training, ~1.1 M parameters	$O(n \cdot k^2 \cdot f)$
Singh et al. [29]	RadioML 2016.10b (160,000 samples, 10 modulations, SNR -20 to +18 dB)	Kera, TensorFlow	CNN (6 conv + 2 dense layers)	~18 min, 336,396 parameters	$O(n \cdot k^2 \cdot f)$
Proposed work	10 MCM types, 10,000 samples (1000 per SNR), SNR -20 to 30 dB	AMD Ryzen 75,700 U CPU (1.80 GHz), Radeon Graphics, 16 GB RAM, Python, and Google Colab TensorFlow, Keras	GTCC + Improved-Inception	0.07 + 0.1286 = 0.1986 inference time (in seconds) per image	$O(n \cdot k^2 \cdot f)$
			GTCC + Improved-Xception	0.07 + 0.1077 = 0.1777 inference time (in seconds) per image	
			GTCC + Improved-ResNet152V2	0.07 + 0.1048 = 0.1748 inference time (in seconds) per image	

**FIGURE 9** | Comparison graph of Improved-InceptionV3, Improved-Xception, and Improved-ResNet152V2 models.**TABLE 10** | Accuracy table of three different models.

Improved transfer learning models	-16 dB	-12 dB	-8 dB	-4 dB	0 dB	4 dB	8 dB	12 dB	16 dB
InceptionV3	88%	88%	93%	96%	97%	97%	97%	97%	98%
Xception	86%	94%	95%	97%	97%	97%	96%	97%	97%
ResNet152V2	96%	96%	98%	99%	99%	100%	100%	100%	100%

GTCC spectrograms to enhance robustness under low SNR conditions. Additionally, improved versions of transfer learning models (InceptionV3, Xception, and ResNet152V2) are employed to boost performance. Despite differences in setup, the proposed method demonstrates superior accuracy,

particularly at SNRs below -10 dB, where existing works are untested. Hence, it can be considered optimal within the evaluated context due to its enhanced noise robustness and classification accuracy. The proposed work is optimal because of the following:

1. GTCC spectrograms outperform STFT in low SNR regimes by leveraging auditory-inspired noise robustness.
2. Improved transfer learning models (e.g., ResNet152V2) achieve state-of-the-art accuracy (100% at  $\geq 4$  dB).
3. Broader SNR coverage ( $-16$  to  $16$  dB) validates robustness in real-world 5G scenarios.

4. Direct improvements over Tailor et al. [5] on the same dataset confirm superiority.

While experimental setups differ, the proposed method's architectural advancements and targeted focus on low SNR conditions justify its optimality (Table 11).

**TABLE 11** | Comparative analysis.

<b>Work</b>	<b>Dataset</b>	<b>Techniques</b>		<b>Accuracy</b>
		<b>Front-end</b>	<b>Back-end</b>	
Hauser et al. [30]	OFDM-QAM, FBMC-OQAM, and UFMC with multi-amplitude gains in noisy channel	PCA	CNN	Average 97.4% within SNR ranging $-5$ to $20$ dB
Wu et al. [4]	4FSK, 2FSK, 4PSK, 2PSK, 4ASK, and 2ASK	High-order cumulants	Deep neural network	99% at $-2$ dB and 100% at $-5$ dB
Jdid et al. [6]	CW, LFM, NLFM, BFSK, BPSK, QPSK, and some combination of them, total 16 types	STFT spectrograms	CNN	99%
Chakravarty and Dua [17]	5G waveform dataset of SNR values $-16, -10, 0$ , and $10$ dB	STFT RGB spectrograms	Xception	$-10$ dB—91% $0$ dB—95% $4$ dB—95% $10$ dB—99%
Proposed work	5G waveform dataset of SNR values $-16, -12, -8, 0, 4, 8, 12$ , and $16$ dB	GTCC spectrograms	Improved-InceptionV3	<b>SNR</b> $-16$ dB 88% $-12$ dB 88% $-8$ dB 93% $-4$ dB 96% $0$ dB 97% $4$ dB 97% $8$ dB 97% $12$ dB 97% $16$ dB 98%
			Improved-Xception	$-16$ dB 86% $-12$ dB 94% $-8$ dB 95% $-4$ dB 97% $0$ dB 97% $4$ dB 97% $8$ dB 96% $12$ dB 97% $16$ dB 97%
			Improved-ResNet152V2	$-16$ dB 96% $-12$ dB 96% $-8$ dB 98% $-4$ dB 99% $0$ dB 99% $4$ dB 100% $8$ dB 100% $12$ dB 100% $16$ dB 100%

## 7 | Conclusion and Future Work

This paper demonstrates that GTCC spectrograms paired with enhanced transfer learning models achieve robust AMC in multicarrier systems under varying SNR conditions. The GTCC front-end, inspired by auditory perception, outperforms conventional feature extraction methods by extracting discriminative features even at extremely low SNR. ResNet152V2 excels among evaluated architectures, delivering 96% accuracy at  $-16\text{dB}$  and near-perfect classification (100%) at  $\text{SNR} \geq 4\text{dB}$ , owing to its residual connections that stabilize gradient flow and refine feature learning in noise. The framework generalizes across a broad SNR range, surpassing prior works restricted to narrower or less challenging SNR intervals. The framework further incorporates Grad-CAM-based visual analysis and testing on colored noisy spectrograms of MCM signals, reinforcing its robustness and interpretability. These results validate the synergy of biologically inspired feature extraction and deep transfer learning for scalable AMC in 5G/6G systems.

Future efforts should explore lightweight architectures (e.g., MobileNet and Vision Transformers) to optimize edge deployment. Hybrid feature extraction methods, such as wavelet-GTCC fusion, could enhance resilience to nonstationary noise. Expanding datasets to include emerging waveforms (OTFS and FBMC) and adversarial channel effects (multipath fading and interference) would improve generalizability. Real-time validation on FPGA/SDR platforms and cross-domain adaptation for non-5G signals (e.g., radar and satellite) remain critical for practical deployment in heterogeneous networks.

### Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

### References

1. K. M. Ho, C. Vaz, and D. G. Daut, “A Wavelet-Based Method for Classification of Binary Digitally Modulated Signals,” in *2009 IEEE Sarnoff Symposium* (IEEE, 2009), 1–5.
2. W. Xie, S. Hu, C. Yu, P. Zhu, X. Peng, and J. Ouyang, “Deep Learning in Digital Modulation Recognition Using High Order Cumulants,” *IEEE Access* 7 (2019): 63760–63766.
3. E. Moser, M. K. Moran, E. Hillen, D. Li, and Z. Wu, “Automatic Modulation Classification via Instantaneous Features,” in *2015 National Aerospace and Electronics Conference (NAECON)* (IEEE, 2015), 218–223.
4. J. Wu, Y. Zhong, and A. Chen, “Radio Modulation Classification Using STFT Spectrogram and CNN,” in *2021 7th International Conference on Computer and Communications (ICCC)* (IEEE, 2021), 178–182.
5. A. Tailor, M. Dua, and P. Verma, “Automatic Classification of Multi-Carrier Modulation Signal Using STFT Spectrogram and Deep CNN,” *Physica Scripta* 99, no. 7 (2024): 076009.
6. B. Jdid, K. Hassan, I. Dayoub, W. H. Lim, and M. Mokayef, “Machine Learning Based Automatic Modulation Recognition for Wireless Communications: A Comprehensive Survey,” *IEEE Access* 9 (2021): 57851–57873.
7. T. J. O’Shea, J. Corgan, and T. C. Clancy, “Convolutional Radio Modulation Recognition Networks,” in *Engineering Applications of Neural Networks: 17th International Conference, EANN 2016* (September 2–5, 2016, Proceedings 17: Springer International Publishing, 2016), 213–226.
8. S. Rajendran, W. Meert, D. Giustiniano, V. Lenders, and S. Pollin, “Deep Learning Models for Wireless Signal Classification With Distributed Low-Cost Spectrum Sensors,” *IEEE Transactions on Cognitive Communications and Networking* 4, no. 3 (2018): 433–445.
9. S. Duan, K. Chen, X. Yu, and M. Qian, “Automatic Multicarrier Waveform Classification via PCA and Convolutional Neural Networks,” *IEEE Access* 6 (2018): 51365–51373.
10. L. Guo, Y. Wang, Y. Liu, Y. Lin, H. Zhao, and G. Gui, “Ultra Lite Convolutional Neural Network for Automatic Modulation Classification in Internet of Unmanned Aerial Vehicles,” *IEEE Internet of Things Journal* 11, no. 11 (2024): 20831–20839.
11. W. Ma, Z. Cai, and C. Wang, “A Transformer and Convolution-Based Learning Framework for Automatic Modulation Classification,” *IEEE Communications Letters* 28 (2024): 1392–1396.
12. H. Ouamna, A. Kharbouche, Z. Madini, and Y. Zouine, “Deep Learning-Assisted Automatic Modulation Classification Using Spectrograms,” *Engineering, Technology & Applied Science Research* 15, no. 1 (2025): 19925–19932.
13. G. Molis, J. Ruseckas, and A. Mackuté-Varoneckienė, “5G Waveforms OFDM FOFDM FBMC UFMC WOLA QAM16&64 [Data set],” Kaggle (2022), <https://doi.org/10.34740/KAGGLE/DSV/3269141>.
14. L. Huang, Y. Zhang, W. Pan, J. Chen, L. P. Qian, and Y. Wu, “Visualizing Deep Learning-Based Radio Modulation Classifier,” *IEEE Transactions on Cognitive Communications and Networking* 7, no. 1 (2020): 47–58.
15. P. Qi, X. Zhou, S. Zheng, and Z. Li, “Automatic Modulation Classification Based on Deep Residual Networks With Multimodal Information,” *IEEE Transactions on Cognitive Communications and Networking* 7, no. 1 (2020): 21–33.
16. N. Chakravarty, M. Dua, and S. Dua, “Automatic Modulation Classification Using Amalgam CNN-LSTM,” in *2023 IEEE Radio and Antenna Days of the Indian Ocean (RADIO)* (IEEE, 2023), 1–2.
17. N. Chakravarty and M. Dua, “Feature Extraction Using GTCC Spectrogram and ResNet50 Based Classification for Audio Spoof Detection,” *International Journal of Speech Technology* 27, no. 1 (2024): 225–237.
18. M. Babu, M. A. Kumar, and S. Santhosh, “Extracting MFCC and GTCC Features for Emotion Recognition From Audio Speech Signals,” *International Journal of Research in Computing, Applications and Robotics* 2, no. 8 (2014): 46–63.
19. D. Wang, C. Shi, J. Li, J. Gan, X. Niu, and L. Xiong, “M-GFCC: Audio Copy-Move Forgery Detection Algorithm Based on Fused Features of MFCC and GFCC,” in *International Artificial Intelligence Conference* (Springer Nature Singapore, 2023), 220–234.
20. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, Rethinking the Inception Architecture for Computer Vision, (2016), In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Pp. 2818–2826).
21. V. C. Burkappalli and P. C. Patil, “Transfer Learning: Inception-V3 Based Custom Classification Approach for Food Images,” *ICTACT Journal on Image & Video Processing* 11, no. 1 (2020): 2245–2253.
22. N. Jinsakul, C. F. Tsai, C. E. Tsai, and P. Wu, “Enhancement of Deep Learning in Image Classification Performance Using Xception With the Swish Activation Function for Colorectal Polyp Preliminary Screening,” *Mathematics* 7, no. 12 (2019): 1170.
23. N. Rachburee and W. Punlumjeak, “Lotus Species Classification Using Transfer Learning Based on VGG16, ResNet152V2, and MobileNetV2,” *IAES International Journal of Artificial Intelligence (IJ-AI)* 11, no. 4 (2022): 1344.

24. H. Xing, X. Zhang, S. Chang, et al., "Joint Signal Detection and Automatic Modulation Classification via Deep Learning," *IEEE Transactions on Wireless Communications* 23 (2024): 17129–17142.

25. Q. Zheng, X. Tian, L. Yu, A. Elhanashi, and S. Saponara, "Recent Advances in Automatic Modulation Classification Technology: Methods, Results, and Prospects," *International Journal of Intelligent Systems* 2025, no. 1 (2025): 4067323.

26. Y. Hu, C. Li, X. Wang, L. Liu, and Y. Xu, "Modulation Recognition of Optical and Electromagnetic Waves Based on Transfer Learning," *Optik* 291 (2023): 171359.

27. S. Pratschner, B. Tahir, L. Marijanovic, et al., Versatile Mobile Communications Simulation: The Vienna 5G Link Level Simulator, (2018), arXiv:1806.03929.

28. Technical Specification Group Radio Access Network, High Speed-Downlink Packet Access: UE Radio Transmission and Reception, May (2002), 3rdGeneration Partnership Project (3GPP) Std. TR 25.890.

29. B. Singh, P. Jindal, P. Verma, V. Sharma, and C. Prakash, "Automatic Modulation Recognition Using Modified Convolutional Neural Network," in *2025 3rd International Conference on Device Intelligence, Computing and Communication Technologies (DICCT)* (IEEE, 2025), 547–550.

30. S. C. Hauser, W. C. Headley, and A. J. Michaels, "October. Signal Detection Effects on Deep Neural Networks Utilizing Raw IQ for Modulation Classification," in *MILCOM 2017–2017 IEEE Military Communications Conference (MILCOM)* (IEEE, 2017), 121–127.