

Review

Not peer-reviewed version

Deep Learning for Automatic Modulation Classification: A Review

[AnuraagChandra Singh Thakur](#)* and [Masudul Imtiaz](#)

Posted Date: 7 April 2026

doi: 10.20944/preprints202604.0446.v1

Keywords: automatic modulation classification; convolutional neural networks; deep learning; rf machine learning; spectrum monitoring; transformers



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

Deep Learning for Automatic Modulation Classification: A Review

AnuraagChandra Singh Thakur * and Masudul Imtiaz

Department of Electrical and Computer Engineering, Clarkson University, Potsdam, NY 13699, USA

* Correspondence: anuthak@clarkson.edu

Abstract

Automatic modulation classification (AMC) is a key component of spectrum awareness, cognitive radio, and signal intelligence, enabling receivers to identify modulation schemes from noisy in-phase and quadrature (IQ) observations. Traditional approaches rely on likelihood-based methods or handcrafted feature extraction, which often struggle under channel impairments and real-world variability. Recent advances in deep learning enable models to learn directly from multiple signal representations, including raw IQ samples, engineered features, and time–frequency or constellation-based encodings, improving adaptability across diverse signal conditions. This paper presents a structured review of deep learning approaches for AMC, including CNNs, RNN/LSTM models, and transformer-based architectures, with a focus on performance, robustness, and system-level trade-offs. We analyze how representation choices, dataset design, and evaluation protocols influence reported results, and highlight key challenges such as domain shift, low-SNR environments, and multi-signal interference. Finally, we outline future directions focused on improving generalization, integrating classical signal processing with learning-based methods, and enabling efficient deployment in real-world and resource-constrained systems.

Keywords: automatic modulation classification; convolutional neural networks; deep learning; rf machine learning; spectrum monitoring; transformers

1. Introduction

Automatic modulation classification (AMC) focuses on identifying the modulation format directly from received signal observations, without relying on prior knowledge of the transmitter. Beyond the modulation label itself, operational systems often seek to infer higher-layer protocol characteristics (e.g., access schemes, coding, or adaptive-rate behavior), making AMC part of a broader “signal intelligence” pipeline. For example, low-power IoT systems such as LoRaWAN adapt parameters (e.g., spreading factor and data rate) to changing link budgets, so a monitoring receiver faces diverse waveforms and operating points even within a single deployment [1–3].

Similarly, spectrum monitoring sites and cognitive radios must operate under uncertain channels and non-stationary interference, motivating methods that scale beyond handcrafted feature extraction.

Classical AMC approaches are typically categorized as likelihood-based or feature-based. Likelihood-based methods compare hypotheses under assumed signal and noise models, but can be computationally expensive and fragile to model mismatch [4,5]. One advantage of CNNs and transformer-based models is that they eliminate the need for manual feature extraction, which is often limited by the designer’s statistical tools and prior domain knowledge. Instead, these models learn hierarchical and task-relevant representations directly from raw data, enabling them to capture complex, nonlinear patterns that may be difficult to model explicitly. This data-driven approach allows for greater adaptability to diverse signal conditions, including varying SNRs, channel impairments, and modulation types, leveraging the power of automatic learning to discover complex patterns and relationships that may not be apparent through traditional, hand-engineered approaches.

Feature-based methods compute descriptors (e.g., higher-order statistics, cyclostationary features) and then apply an AMC classifier. Broad surveys summarize these approaches and their limitations in non-ideal channels and at scale [6]. More recently, online and kernel-based approaches have explored alternative classification pipelines [7]. Deep learning (DL) methods, by contrast, can learn representations directly from data and have become a major direction in RF machine learning (RFML), particularly after the “convolutional radio” paradigm showed that compact CNNs can learn discriminative features from raw IQ samples [8,9]. As the field of DL has advanced, the range of tools used in AMC has expanded as well. These tools include convolutional neural networks (CNNs), Transformers, LSTMs, and RNNs, all of which will be discussed in this paper.

Beyond modulation recognition, similar DL approaches have also been applied to detect Forward Error Correction (FEC) coding schemes in non-cooperative communication environments where the receiver has no prior knowledge of the transmitter configuration. For example, Ramabadran et al. proposed a method for blind recognition of LDPC code parameters under erroneous channel conditions by exploiting the structural properties of sparse parity-check matrices to infer coding parameters from noisy observations [10]. More recently, DL methods have been introduced to improve robustness in low signal-to-noise environments. Zhang and Zhang proposed a cascade neural network architecture that combines denoising and classification networks to automatically learn LDPC (Low-Density Parity Check) code features from received signals, improving blind recognition performance under noisy conditions [11]. It employs a transformer-based method to accomplish this. Similarly, CNNs have been applied to communication signal classification tasks due to their ability to learn spatial correlations in structured signal representations. Yan et al. demonstrated that CNN-based models can effectively recognize coding structures such as space-time block codes by extracting discriminative features from received signal matrices [12].

Table 1. Extended RF Inference Tasks Beyond Modulation Classification.

Modulation Classification	Identifying the modulation scheme (e.g., BPSK, QPSK, QAM) from received signals; core task in AMC.	CNNs, LSTMs, transformers [9,47,48]
FEC / Coding Scheme Recognition	Inferring forward error correction schemes (e.g., LDPC, convolutional codes) from received data.	CNN and DL-based LDPC recognition [10,11]
Interference Classification	Identifying types of interference (e.g., WiFi vs Bluetooth vs noise sources).	Semi-supervised CNNs [37]
Open-Set / Unknown Signal Detection	Detecting signals not seen during training and rejecting unknown classes.	Domain generalization and OTA evaluation [16,17]
Channel / Impairment Estimation	Estimating channel effects such as fading, frequency offset, phase noise, and hardware distortions.	DL-based correction and estimation [18,21]
Signal Detection / Spectrum Sensing	Determining the presence or absence of signals in a frequency band under noise and interference.	DL-based spectrum monitoring [22,32]
Multi-Signal / Emitter Separation	Identifying and separating overlapping signals from multiple transmitters in time/frequency.	CNN-based and radar-signal approaches [13,14]
Radio Access Technology (RAT) Identification	Detecting communication protocols (e.g., LTE, WiFi, LoRa) using time–frequency or spectral patterns.	CNN on spectrograms [32,33]
Protocol Behavior / Adaptation Detection	Characterizing adaptive behaviors such as rate control, power adjustment, or IoT link adaptation.	IoT and LoRa-based analysis [1,2]
Device / Transmitter Identification (RF Fingerprinting)	Identifying unique hardware signatures of transmitters based on RF impairments.	DL-based RF fingerprinting [19,20]

A significant amount of literature exists on DL techniques for FEC detection in addition to AMC. The works discussed here are not exhaustive but rather provide a brief overview to illustrate representative approaches and current research trends. DL tools can be applied throughout the entire signal intelligence pipeline, from modulation and forward error correction (FEC) to protocol detection, making this an important area for research and awareness. This paper presents Table 3, which provides an overview of these tasks and representative approaches beyond AMC.

2. Review Strategy

This review synthesizes DL-based AMC research, with an emphasis on understanding these dependencies and clarifying where performance claims are effectively translated, or fail to be translated, to operational RF environments. The paper identifies key limitations of current RFML techniques as well as outlines promising directions for future research. In this paper, we survey five major areas of progress:

1. **Signal Representations and Dataset Design:** Analysis of raw IQ data, engineered features, and time–frequency or constellation-based representations, along with the impact of dataset realism and evaluation protocols on reported performance.
2. **Convolutional Neural Networks (CNNs):** CNN-based approaches for AMC, including baseline architectures, robustness enhancements, hybrid feature integrations, and lightweight models for deployment.
3. **Sequence Models (RNNs and LSTMs):** Temporal modeling techniques that capture sequential dependencies in RF signals, with discussion of their strengths, limitations, and role relative to CNN-based methods.
4. **Transformer-Based Architectures:** Attention-driven models that enable global context learning from IQ sequences and derived representations, including design considerations and performance trade-offs.
5. **Real-World Challenges and System-Level Considerations:** Challenges include domain shift, low-SNR environments, multi-signal interference, and deployment constraints, alongside methods such as transfer learning and ensemble learning to improve accuracy.

2.1. Review Methodology

This review synthesizes advances in deep learning-based automatic modulation classification (AMC), with an emphasis on methods that operate on raw IQ data and related signal representations. Rather than restricting the survey to a fixed time window, we focus on works that are most relevant to current research trends and practical deployment considerations.

Information Sources and Search Strategy. We drew from major digital libraries including IEEE Xplore, ACM Digital Library, arXiv, and Google Scholar. Search queries combined AMC and RF machine learning terminology with deep learning keywords, such as:

“Automatic Modulation Classification”, “RF Machine Learning”, “RF Deep Learning”, “CNN”, “RNN”, and “Transformer”.

These queries were iteratively refined to capture both foundational and recent contributions.

Selection Approach. Rather than applying strict inclusion or exclusion filters, we adopted a relevance-driven approach. Papers were selected based on their contribution to understanding deep learning methods for AMC, including architectural innovations, representation choices, and evaluation strategies. Foundational works were included where necessary to provide context, alongside more recent studies reflecting current trends.

Scope of Reviewed Work. The survey emphasizes studies that explore learning-based modulation classification using raw IQ samples, engineered features, or time–frequency representations. We also include works that highlight practical challenges such as domain shift, low-SNR conditions, and multi-signal environments, even when AMC is part of a broader signal intelligence pipeline.

Organization of the Review. The selected literature is organized into five major categories: signal representations and datasets, CNN-based models, sequence models (RNNs/LSTMs), transformer-based architectures, and system-level challenges. The goal is not to exhaustively enumerate all prior work, but to synthesize key ideas, trends, and trade-offs that shape modern AMC research and deployment. More importantly, it highlights a path forward that avoids recurring dataset-driven biases and promotes the fusion of novel architectures and approaches.

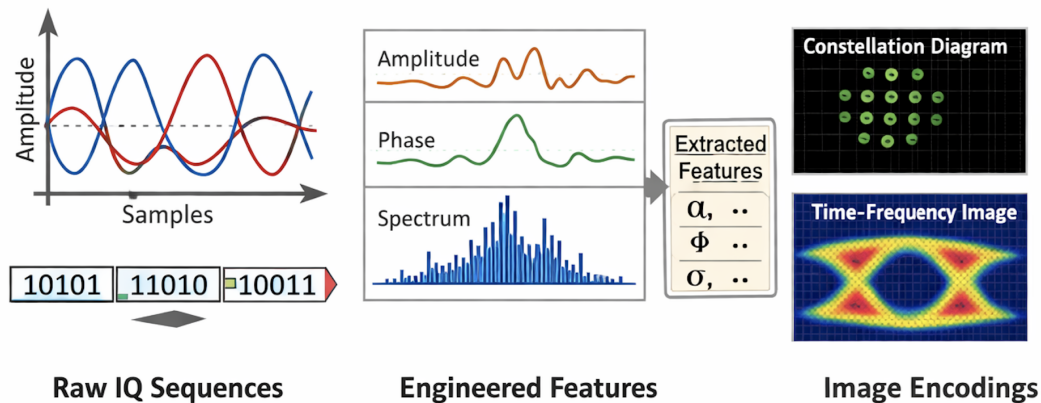


Figure 1. Representation strategies for deep learning-based modulation classification. The figure shows three common signal representations: (Left) Raw IQ sequences, with in-phase (I) and quadrature (Q) samples; (Middle) Engineered features, where transformations such as amplitude/phase and spectral features are extracted; and (Right) Image encodings, where signals are converted into visual forms such as constellation diagrams or time–frequency images.

Table 2. Comparison of Signal Representations for Deep Learning-Based AMC.

Representation	Advantages	Limitations
Raw IQ Sequences	<ul style="list-style-type: none"> • No preprocessing; preserves full amplitude and phase information • Enables true end-to-end learning from raw observations • Avoids bias from handcrafted features • Captures fine-grained temporal dynamics and phase transitions • Suitable for real-time/streaming systems (low latency) • Flexible across signal types and channel conditions • Retains information for joint learning of impairments (noise, drift, offsets) 	<ul style="list-style-type: none"> • Sensitive to noise and channel impairments (low SNR degradation) • Phase/frequency offsets can distort learned patterns • Requires large datasets to learn invariances • Harder to learn robustness without augmentation • Multipath fading and nonlinear effects corrupt structure • Higher training complexity due to limited inductive bias
Engineered Features	<ul style="list-style-type: none"> • Incorporates domain knowledge (e.g., spectral, wavelet features) • Improved robustness to noise and channel effects • Reduces learning complexity via compact representations • More stable across varying SNR conditions • Highlights physically meaningful signal characteristics • Integrates well with classical signal processing pipelines 	<ul style="list-style-type: none"> • Requires expert-driven feature design • Potential information loss during transformation • Limited generalization across datasets/hardware • Adds preprocessing latency and system complexity • Reduced adaptability to unseen modulation types • Possible mismatch with learned representations
Image Encodings	<ul style="list-style-type: none"> • Enables use of vision-based models (CNNs, ViTs) • Provides interpretable representations (spectrograms, constellations) • Highlights spatial modulation patterns • Leverages mature computer vision architectures • Effective in high-SNR conditions • Captures joint time–frequency structure 	<ul style="list-style-type: none"> • Requires preprocessing (e.g., STFT), increasing overhead • Often assumes synchronization (timing, carrier alignment) • Noise degrades constellation separability • High-order modulations become difficult under noise • Multipath and interference distort visual structure • Loss of temporal information in static images • Higher memory and compute requirements

3. Signal Model, Task Definitions, and Evaluation Dimensions

A typical receiver observes a complex baseband sequence $\mathbf{x} \in \mathbb{C}^N$ generated by an unknown modulation $m \in \mathcal{M}$ through an unknown channel and impairments. A common abstraction is

$$x[n] = (s_m[n] * h[n]) e^{j(2\pi\Delta f n + \phi)} + w[n], \quad n = 0, \dots, N-1, \quad (1)$$

where $s_m[n]$ is the transmitted (baseband) signal for modulation m , $h[n]$ is the channel impulse response, Δf and ϕ represent carrier-frequency and phase offsets, and $w[n]$ is additive noise. AMC seeks a classifier f_θ that predicts $\hat{m} = f_\theta(\mathbf{x})$.

3.1. Closed-Set vs. Open-set AMC

Most published benchmarks assume *closed-set* classification: the label set \mathcal{M} is fixed and exhaustive. In practice, receivers can encounter unknown or previously unseen waveforms, mixed radar/communications emissions, and novel protocol variants. Thus, a deployed system benefits from open-set rejection, novelty detection, and the ability to defer decisions when uncertainty is high. Although these aspects are not consistently evaluated, they are increasingly reflected in discussions of “real-life complications,” such as varying signal lengths, different SNR levels, and multi-signal mixtures in the literature [13–15].

3.2. Channel Impairments and Domain Shift

Deep models trained on idealized simulations can be brittle when the test distribution differs, e.g., due to oscillator offsets, multipath profiles, front-end nonlinearities, or different sampling conditions. Work emphasizing *over-the-air* evaluation shows that bridging this gap is a first-order challenge for DL AMC [16,17]. Several lines of research incorporate robustness modules, feature transformations, and multi-channel estimation to reduce sensitivity to such shifts [18,19,21].

3.3. Metrics and Protocols

Most papers report accuracy vs. SNR curves and confusion matrices on RadioML-style datasets (Table I). However, operational systems often face non-uniform SNR distributions, changing SNR within a capture, class imbalance, and mixed signal occupancy. Consequently, evaluations that include cross-SNR generalization, cross-dataset transfer, calibration, and out-of-distribution detection are critical, but less common. Distributed sensing adds further constraints, such as limited on-device compute and intermittent connectivity [22].

4. Datasets and Signal Representations

Benchmark datasets strongly influence architecture choices and reported results. RadioML datasets popularized raw IQ inputs for modulation recognition; multiple generations expand the number of modulation classes, sample length, and channel models. HisarMod2019.1 provides a broader modulation set under controlled conditions and has been used to test modern CNN backbones and ensemble variants. A recurring critique is that public datasets frequently contain *one modulation per capture* with known center frequency and simplified impairment models, which can overstate real-world readiness. Furthermore, many widely used datasets fail to report critical parameters, such as sampling rate, baud rate, and other key signal characteristics, limiting their ability to capture the true diversity of real-world signals.

4.1. Representations

Three primary representation families recur (as depicted in Figure 1):

- **Raw IQ sequences:** used in early end-to-end CNN baselines and later in transformer work.
- **Engineered features:** amplitude/phase sequences, spectral features, or wavelet transforms, sometimes used with sequential models or hybrid CNNs [23,24].
- **Image encodings:** constellation diagrams or time–frequency images (e.g. eye diagrams) that permit reuse of vision CNNs (AlexNet/VGG/ResNet variants) [25].

Image encodings can introduce inductive biases and leverage mature vision architectures, but add preprocessing overhead and can assume prior synchronization or segmentation; raw IQ avoids that overhead but may require augmentation to learn invariances.

These datasets differ in channel realism, modulation diversity, and SNR coverage, with HisarMod2019.1 incorporating more realistic multipath and fading conditions compared to synthetic RadioML datasets.

Table 3. Common AMC datasets referenced in the reviewed literature.

Dataset	# Classes	Sample Shape	SNR Range (dB)
RadioML2016.10A [40]	11	2×128	-20 to 18
RadioML2016.10B [41]	11	2×128	-20 to 18
RadioML2016.04C [40]	11	2×128	-20 to 18
RadioML2018.01A [40]	24	2×1024	-20 to 30
HisarMod2019.1 [42]	26	2×1024	-20 to 18

4.2. Labeling Scope: Modulation only vs. Richer Parameters

Several adjacent tasks motivate extensions beyond modulation labels. Examples include identifying radio access techniques via time–frequency analysis and CNNs, characterizing adaptive behaviors in IoT links, and recognizing coding parameters under channel errors (e.g., LDPC parameter recognition). While this review focuses on modulation recognition, these works highlight that realistic systems often require multi-attribute inference.

5. A Taxonomy of DL-Based AMC Approaches

5.1. CNN-Centric Modulation Recognition

CNNs became a standard baseline for AMC because convolution captures local structure in IQ time series and supports highly parallel inference. The seminal “convolutional radio” work demonstrated that compact CNNs can learn discriminative features from raw IQ and compete strongly with classical baselines on RadioML-style data. Subsequent studies explored deeper CNNs, residual architectures, and alternative input transformations [26,27].

5.2. From Baseline CNNs to Improved Accuracy

A broad set of papers reports CNN variants that improve recognition across SNR or target-specific channel conditions. Examples include cognitive-radio waveform recognition [28], CNNs tailored for complicated channels, and robust VHF modulation recognition [29,30]. Data-driven DL for cognitive radios further emphasizes end-to-end training and scalability [31]. In spectrum monitoring contexts, end-to-end learning from spectrum data has been used for wireless signal identification beyond a narrow modulation set, highlighting the potential for broader label spaces and sensing conditions [32].

5.3. Robustness via Correction and Auxiliary Features

Several works attempt to explicitly address the mismatch between training and deployment. Parameter-estimation-and-transformation models aim to normalize impairments before classification. Learnable distortion correction modules provide end-to-end trainable front-ends that can mitigate distortions without hard-coded signal processing blocks. Blind channel identification features can be fused with deep networks for generalized recognition under unknown channels [33]. Over-the-air evaluation studies emphasize that these robustness ideas should be validated on realistic hardware channels, not only on simulated impairments.

5.4. Alternative Feature Extractions and Hybrid Encodings

Wavelet-assisted CNNs combine time–frequency localization with learned features. Time–frequency analysis paired with CNNs has been used to identify radio access techniques. Constellation-diagram-based pipelines treat AMC as image classification and leverage strong vision inductive biases. Radon-transform-based image features combined with CNNs further extend this direction [34]. The benefit of image-like encodings is access to mature vision backbones; the tradeoff is sensitivity to preprocessing, potential loss of phase continuity, and additional runtime overhead.

5.5. Architectural Variations: Dilation, Dropout, and Fusion

To increase the receptive field without excessive depth, dilated CNNs have been proposed for modulation recognition under low SNR. Regularization choices such as dense-layer dropout have been explored to reduce overfitting and improve generalization [35]. Multi-feature fusion (e.g., combining multiple learned features or representations) has also been reported to improve accuracy. Related work combines multiple neural networks or multi-scale designs to improve performance on larger modulation sets.

5.6. Efficiency and Deployment-Oriented CNNs

Practical spectrum monitoring motivates lighter models and faster inference. Lightweight CNN architectures reduce parameters and inference time while aiming to preserve accuracy on RadioML benchmarks [36]. Semi-supervised CNNs have been used to reduce labeling burden when identifying interference sources, reflecting the broader need for label-efficient learning in realistic monitoring pipelines [37]. Distributed sensing and low-cost deployments also motivate model compression and aggregation strategies across sensors [38,39].

6. Sequence Models: RNN and LSTM Approaches

RNNs and LSTMs explicitly model temporal dependencies and have been applied to AMC on RadioML benchmarks. Early work reported competitive accuracy with recurrent architectures for modulation recognition [43]. LSTM-based networks can, in principle, capture longer-range dependencies than fixed-size convolution kernels, but incur higher computational cost and limited parallelism compared to CNNs and transformers.

A recurring theme is *representation choice*: passing raw IQ alone can be less effective than feeding amplitude/phase or frequency-domain features that align with the recurrent model's temporal inductive bias. In addition, sampling conditions (e.g., samples per symbol) can vary across real-world captures, motivating training strategies that span multiple sampling regimes and heterogeneous sensors. Overall, LSTMs remain useful baselines and can complement CNN features, but transformer models increasingly dominate where training data is abundant and parallel computing is available [44,45].

7. Attention Mechanisms and Transformers

Attention mechanisms model long-range dependencies by learning content-based interactions across sequence elements. The transformer architecture replaces recurrence with self-attention and feedforward blocks [46]. Let $\mathbf{Z} \in \mathbb{R}^{T \times d}$ denote the input token embeddings, where T is the sequence length and d is the embedding dimension. The attention mechanism computes

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V, \quad (2)$$

where $Q = ZW_Q$, $K = ZW_K$, and $V = ZW_V$ are the query, key, and value matrices, respectively. Here, W_Q , W_K , and W_V are learned projection matrices, and d_k denotes the dimensionality of the key vectors. This formulation enables global context modeling with efficient parallel computation, making transformers suitable for long RF sequences and real-time analysis. By adaptively weighting signal segments, they capture long-range dependencies beyond the reach of convolutional or recurrent models.

This structure supports global context aggregation and parallel training (can heavily leverage modern GPU-based parallel computing), making it attractive for long RF captures and real time analysis.

7.1. Transformer-Based AMC

Transformer AMC models typically tokenize IQ sequences (or patches of derived representations), add positional information, and apply multi-head self-attention to capture global context. Reported

results suggest competitive or improved performance relative to CNN/LSTM baselines on RadioML-style datasets, particularly when model capacity and training data are sufficient [47–49].

However, transformers are often *data-hungry*: in low-SNR regimes where discriminative information is weak, accuracy can remain limited without additional augmentation or explicit impairment handling. The broader transformer literature suggests that pretraining and scale can compensate for data hunger; an open question is how best to translate these ideas to RFML under constrained labeled data. Furthermore, a spectrogram, which captures both time and frequency information, can be represented as an image and processed using transformer-based architectures such as vision transformers (ViT) [50].

Transformer models are not a single, uniform architecture but a flexible framework with many design variations that significantly affect performance. Key design choices include how input data is segmented into tokens, the number of tokens used to represent a signal or sequence, and the degree of overlap between tokens. For example, in signal processing tasks like AMC, one can divide IQ data into fixed-length segments (tokens), and optionally allow overlap between adjacent segments to preserve temporal continuity. The size of each token determines the granularity of local features captured, while the total number of tokens influences the model's ability to learn long-range dependencies. These choices introduce trade-offs between computational cost, resolution, and contextual awareness, making the tokenization strategy a critical component of transformer-based models.

In addition to tokenization strategies, there are many transformer variants tailored to different applications and learning paradigms. Models such as BERT (Bidirectional Encoder Representations from Transformers) focus on bidirectional context and are widely used for representation learning, while others like GPT emphasize autoregressive generation [51,52]. In engineering and signal-processing contexts, transformers may be adapted with custom embeddings, positional encodings, or hybrid architectures that combine convolutional layers with attention mechanisms. This diversity highlights that transformer models should be viewed as a broad family of architectures rather than a single fixed design, with performance heavily dependent on how the model is configured for the specific data and task.

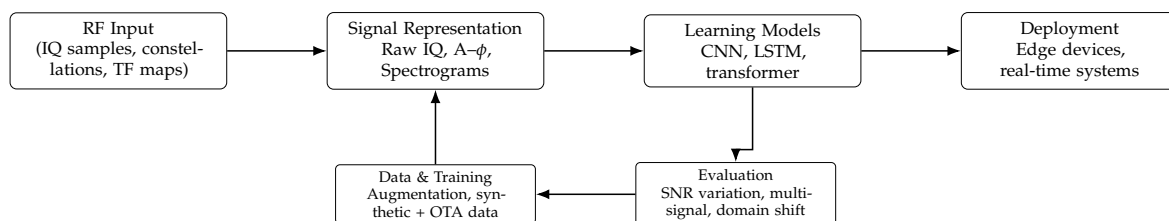


Figure 2. System-level view of deep learning-based automatic modulation classification (AMC). The pipeline highlights the progression from RF inputs to signal representations and model inference, along with training and evaluation under realistic conditions such as varying SNR, multi-signal environments, and domain shift.

7.2. Large Language Model Meta AI for AMC Adaption in Realistic Environment

LLaMA (Large Language Model Meta AI) illustrates how transformer architectures can scale through large-scale pretraining to learn transferable representations. While developed for natural language processing, it shares the same self-attention foundation as transformer models used in RF machine learning, highlighting the domain-agnostic nature of these architectures.

In AMC, similar pretraining strategies can improve generalization across channel conditions, SNR levels, multi-signal environments, and hardware impairments. By leveraging diverse signal datasets, transformer-based models can learn robust representations that transfer to tasks such as modulation classification and multi-signal recognition. As a result, LLaMA serves as a concrete example of the broader transformer framework's flexibility and scalability, reinforcing the idea that transformer architectures are not limited to a single domain but can be extended to complex signal analysis problems [53].

7.3. Time–Frequency Attention and Hybrid CNN-Attention Designs

Rather than replacing convolution, some work adds attention to CNN backbones to emphasize discriminative regions in time–frequency space [54].

Hybrid designs aim to combine CNN locality (robust local feature extraction) with attention’s ability to relate distant parts of a capture. These hybrids can reduce parameter count while improving recognition under varying SNR and class sets. This approach is particularly well-suited for modulation detection, given the wide range of modulation schemes, each exhibiting distinct local and global features [55].

7.4. IoT- and Systems-Oriented Transformer Variants

IoT monitoring introduces constraints (limited power/compute, bursty transmissions, heterogeneous devices). Transformer variants have been proposed to enhance recognition in IoT-oriented AMC settings, complementing earlier work on low-cost distributed sensors for signal classification [22,32,48].

Binary transformers were used to detect a signal of interest within multi-signal, mixed-frequency environments. Unlike conventional transformer models that are trained on large multi-class datasets (e.g., 24 modulation classes) simultaneously, this approach adopts a modular design, where lightweight transformer models are trained for a single signal of interest. This modular strategy improves efficiency, scalability, and deployment in resource-constrained settings. Additionally, transfer learning was explored, enabling continuous training and adaptation, making this approach well-suited for evolving signal environments. These systems-level settings and novel approaches motivate research that co-designs models, sensing hardware (receiver and transmitter), and communication protocols [56].

8. Ensembles, Fusion, and Distributed Spectrum Monitoring

Ensemble learning combines multiple base learners to reduce variance and mitigate local minima; for AMC, ensembles can average predictions across architectures or across SNR-specialized experts. Ensembles of CNN backbones have been reported to improve accuracy on HisarMod2019.1 and related datasets [57]. Fusion strategies for CNN-based AMC include feature-level and decision-level fusion, often trading computational cost for accuracy gains [58–60]. Combining multiple neural networks has also been studied as a way to improve generalization across modulation sets and SNR conditions [61].

Distributed monitoring settings introduce additional dimensions: low-cost sensors produce heterogeneous data quality; inference may be split between edge devices and a coordinator; and models may be trained or updated in a distributed manner.

These settings align AMC more closely with federated/distributed learning and systems-level constraints than with single-device classification.

9. Transfer Learning for AMC

Domain adaptation is critical in AMC, as models trained on one signal distribution often fail under different channel, noise, or hardware conditions. By aligning feature representations across domains, these methods improve generalization to real-world environments. Recent work has explored the use of transfer learning to improve automatic AMC under challenging wireless channel conditions. One study proposes a model that combines convolutional neural networks (CNNs) with a self-attention mechanism, allowing the network to better capture both local signal features and long-range dependencies within wireless signal representations. This architecture improves feature extraction and classification performance by leveraging transfer learning to reuse knowledge learned from related datasets. Another study investigates transfer learning as a tool for guiding noise reduction in AMC systems, where knowledge from pre-trained models is used to improve signal denoising and classification performance in low signal-to-noise ratio environments. Together, these approaches demonstrate how transfer learning can enhance robustness, improve feature representation, and increase classification accuracy for wireless signal recognition tasks [54,62].

10. Beyond Communications-Only AMC: Radar and Multi-Signal Mixtures

While many benchmarks focus on communications waveforms, radar emissions (e.g., LFM chirps) and mixed radar–communications spectra are common in practice [13,32]. Radar signal modulation recognition has been explored via fusion-image feature extraction [58–60].

More recent work considers mixed datasets that include both radar and conventional communications modulations and explores CNN-based recognition under multi-component conditions [14].

These studies motivate more realistic multi-emitter benchmarks and algorithms that can disentangle overlapping signals rather than assuming a single dominant emitter.

11. Practicality and Sustainability

Beyond the model family, several practical factors repeatedly determine whether published gains translate to new settings.

Table 4. Advantages and disadvantages of DL network models.

DL Model	Advantages	Disadvantages
CNN	Learns local features and translation invariance; efficient and highly parallelizable; robust to small shifts and noise	Requires fixed-size inputs; limited ability to model long-range dependencies; may need deeper architectures or large receptive fields
RNN	Captures temporal dependencies; processes sequential data; suitable for time-series modeling	Suffers from vanishing and exploding gradients; limited long-term memory; slow training due to sequential processing
LSTM	Handles long-term dependencies; improved gradient flow via gating mechanisms; more stable than vanilla RNNs	Higher computational complexity; slower training; less parallelizable than CNNs and transformers
Transformer	Models long-range dependencies using self-attention; highly parallelizable; captures global context effectively	High computational and memory cost; data-hungry; sensitive to tokenization and architectural design

11.1. Training Data Volume, Augmentation, and Synthetic-to-Real Gaps

Synthetic datasets enable controlled sweeps over SNR and modulation classes, but the learned decision boundary can encode simulator-specific artifacts rather than robust modulation cues. Over-the-air studies show that training on realistic hardware and channel effects is important, and that models can degrade when those effects are absent from training. A common mitigation is to broaden training distributions via augmentation (frequency/phase offsets, timing jitter, fading models) and to incorporate correction modules that are optimized jointly with the classifier. Data generation strategies for CNN-based AMC further highlight that the *way* training examples are produced (e.g., image-like encodings and scaling choices) can affect generalization [63].

11.2. Model Selection Under Latency and Memory Constraints

Spectrum monitoring and embedded receivers impose latency and power budgets. Lightweight CNNs target these constraints directly by reducing parameters and inference time. Ensembles and fusion methods often improve accuracy but can increase compute and memory footprint. Transformer-based models offer parallelism and global context, but can be expensive without careful tokenization and architecture scaling. For distributed sensor networks, systems-level considerations such as communication cost and heterogeneous data quality become first-order design constraints.

11.3. Benchmarking Pitfalls and a Minimal Reporting Checklist

Published AMC results often vary significantly due to subtle differences in dataset splits, preprocessing pipelines, and class or SNR balancing strategies. These variations can lead to inconsistent and difficult-to-compare results across studies.

In practical wireless environments, signals are rarely isolated. Wideband spectrum captures often contain multiple coexisting transmissions that overlap in time and frequency. As a result, the common assumption of a single signal of interest (SOI) does not hold in realistic deployments. Instead, signal detection typically begins with transforming raw IQ data into time–frequency representations (e.g., spectrograms) to localize energy in both domains.

In such representations, bursts of energy above the noise floor are often treated as candidate signals of interest. However, this process is highly sensitive to preprocessing choices such as FFT size, window length, and overlap. If these parameters are not carefully selected, weaker signals or short-duration transmissions may remain undetected, particularly under low SNR conditions. This introduces a critical source of bias, where models are evaluated only on clearly visible signals while missing more challenging cases.

Furthermore, modern architectures such as transformers and vision-based models rely on structured representations like spectrograms, requiring effective extraction of relevant time–frequency features. In multi-signal environments, overlapping transmissions can lead to interference, feature entanglement, and degraded separability between modulation classes. This makes multi-signal interference one of the most significant and under-addressed pitfalls in current AMC research.

Finally, SNR remains a fundamental limitation. Signals below the noise floor may not be reliably detected or represented in preprocessing stages, preventing downstream models from learning meaningful features. Consequently, benchmarking that does not explicitly account for low-SNR conditions or detection failures may overestimate real-world performance.

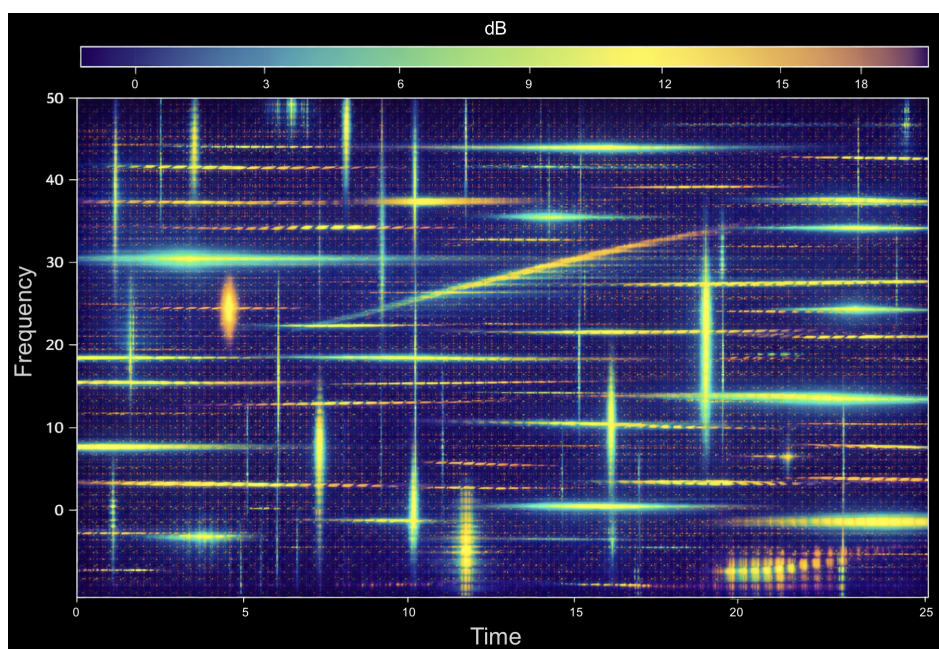


Figure 3. Complex spectrogram demonstrating overlapping transmissions, interference, and the difficulty of detecting weak signals in realistic RF environments.

Table 5. Benchmarking checklist for DL-based AMC studies (summary of recurring pitfalls).

Pitfall	Impact	What to report
Unclear dataset versioning	Results are difficult to reproduce and compare across studies	Dataset version, number of classes, sample length, and generation procedure
SNR imbalance	Inflated or misleading accuracy due to dominance of high-SNR samples	Train/test SNR distribution and per-SNR accuracy curves
Preprocessing variability (TF, constellations)	Hidden biases and inconsistent feature representations	Exact preprocessing pipeline and parameters (e.g., STFT window, normalization)
No cross-domain testing	Unknown robustness under real-world deployment conditions	Cross-dataset evaluation and over-the-air (OTA) testing
Ignoring calibration uncertainty	Overconfident predictions and unreliable decision thresholds	Calibration method (e.g., temperature scaling) and rejection criteria
Single-signal assumption	Unrealistic performance in environments with overlapping signals	Evaluation on multi-signal or interference scenarios
Synthetic-only training data	Poor generalization to real-world RF due to domain shift	Use of OTA data or domain adaptation techniques
Lack of channel variability	Models fail under different propagation conditions (fading, Doppler, offsets)	Channel models used and impairment settings
Signal parameter variation (not modeled)	Performance degradation due to unmodeled variations in signal parameters	Baud rate, sampling rate, sideband effects, RF measurement errors, and front-end components (modulator/demodulator, imaging artifacts)
Fixed input assumptions	Reduced robustness to varying sampling rates, symbol lengths, or bandwidths	Input configuration and preprocessing strategy
Ignoring computational constraints	Models may be impractical for real-time or edge deployment	Model size, FLOPs, inference latency, and hardware setup
Limited evaluation metrics	Accuracy alone does not reflect real-world performance	Additional metrics (precision/recall, confusion matrix, SNR-wise results)
Class imbalance	Bias toward dominant modulation classes	Class distribution and balancing strategy

12. Open Problems and Research Directions

12.1. Robustness and Domain Generalization

Generalization across hardware, channel statistics, sampling rates, and carrier offsets remains a central barrier. Promising directions include learning invariances via augmentation, explicit impairment modeling, self-supervised pretraining on unlabeled RF, and modular front-ends that perform correction jointly with classification. Additionally, online multi-feature classification schemes [64] and semi-supervised learning [37] can help cope with non-stationary environments and scarce labels.

12.2. Low-SNR Recognition and Impairment-Aware Learning

Low SNR remains difficult across model families: even high-capacity transformers can struggle when modulation signatures are swamped by noise. Approaches that incorporate denoising, equalization, or impairment-aware objectives may be necessary, but raise questions about required side information and whether such blocks should be learned or engineered.

12.3. Beyond Single-Signal, Closed-Set Classification

Many public benchmarks assume a single active emitter and a fixed label set. However, real spectra exhibit multiple overlapping signals, intermittent occupancy, and unknown classes. Recent work begins to model multi-signal compositions and radar–communication mixtures, but broader progress will require datasets and metrics that reflect open-set and multi-emitter realities [14].

12.4. Data Realism, Reporting Standards, and Reproducibility

Reported accuracy is often sensitive to train/test splits, SNR balancing, and preprocessing. Community progress would benefit from standardized reporting: dataset versioning, impairment models, calibration, robustness tests, and cross-dataset transfer evaluations. Surveys of DL AMC applications emphasize the importance of such standards as the field matures. The large volume of available data also makes semi-supervised learning approaches particularly suitable [37].

12.5. Edge Deployment and Distributed Sensing

Spectrum monitoring at scale demands low-latency, low-power inference. Lightweight architectures [36] and distributed learning approaches are steps toward practical deployment, but remain constrained by label scarcity, communication limits, and heterogeneous sensor quality. Research that co-optimizes sensing, labeling, and inference under resource constraints remains critical.

12.6. Integrating DL with Classical Signal Processing

There exists substantial domain knowledge in classical signal processing that remains under-explored in DL-based RF classification systems. Established techniques such as filtering, Fourier transforms, and discrete wavelet transforms provide structured representations of spectral and temporal characteristics [19,20,23,32]. Additionally, advanced tools such as polyphase channelization enable selective analysis of signals within specific frequency sub-bands, offering improved interpretability and spectral isolation [65]. Integrating these principled signal processing methods with modern learning architectures presents a practical and promising direction for future research [66]. This direction aligns with the broader paradigm of physics-aware machine learning model design, where domain-specific signal processing knowledge is systematically integrated into deep learning architectures.

13. Conclusions

Deep Learning has reshaped AMC by enabling end-to-end learning from raw or lightly processed RF observations. CNNs remain strong, efficient baselines; LSTMs provide sequence modeling but can be compute-intensive; attention and transformer models offer global context and scalability but often require more data and careful training. Across architectures, dataset realism and evaluation methodology strongly govern conclusions. Future advances are likely to come from (i) more realistic and diverse datasets, (ii) robustness and domain generalization methods validated over-the-air, and (iii) algorithms and systems that explicitly handle multi-signal and open-set conditions under edge constraints. In addition, there remains substantial untapped potential in integrating classical signal processing with deep learning-based RF classification systems. Established techniques such as filtering, Fourier transforms, and discrete wavelet transforms provide structured representations of spectral and temporal characteristics [19,20,23,32]. Advanced methods such as polyphase channelization further enable selective analysis of signals within specific frequency sub-bands, improving interpretability and spectral isolation [65]. Bridging these principled signal processing approaches with modern learning architectures represents a practical and promising direction for improving robustness, efficiency, and real-world applicability [66].

Author Contributions: Please complete during submission. Suggested roles: Conceptualization, A.C.S.T. and M.I.; methodology, A.C.S.T.; software, A.C.S.T.; writing—original draft, A.C.S.T.; writing—review and editing, A.C.S.T. and M.I.; supervision, M.I.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Aldhaheri, Lameya & Alshehhi, Noor & Manzil, Irfana & Khalil, Ruhul Amin & Javaid, Shumaila & Saeed, Nasir & Alouini, Mohamed-Slim. (2024). LoRa Communication for Agriculture 4.0: Opportunities, Challenges, and Future Directions. 10.48550/arXiv.2409.11200.
2. Kufakunesu, R.; Hancke, G.P.; Abu-Mahfouz, A.M. A Survey on Adaptive Data Rate Optimization in LoRaWAN: Recent Solutions and Major Challenges. *Sensors* 2020, 20, 5044. <https://doi.org/10.3390/s20185044>
3. A. S. Thakur and M. H. Imtiaz, "Long Range (LoRa) Agriculture Network in Northern New York: A Scoping Review," 2025 IEEE 16th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), Yorktown Heights, NY, USA, 2025, pp. 0374-0381, doi: 10.1109/UEMCON67449.2025.11267659.
4. Bahloul, Mohammad Rida et al. "An Efficient Likelihood-Based Modulation Classification Algorithm for MIMO Systems." *ArXiv abs/1605.07505* (2016): n. pag.
5. Jin, X., Zhou, X. A new likelihood-based modulation classification algorithm using MCMC. *J. Electron.(China)* 29, 17–22 (2012). <https://doi.org/10.1007/s11767-012-0772-4>
6. O. Dobre, "Survey of Automatic Modulation Classification Techniques: Classical Approaches and New Trends," *IET Communications*, 2007, doi: 10.1049/IET-COM:20050176.
7. X. Li, Z. Jiang, K. Ting, and Y. Zhu, "An Online Automatic Modulation Classification Scheme Based on Isolation Distributional Kernel," *arXiv:2410.02750*, 2024.
8. Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation Applied to Handwritten Zip Code Recognition," *Neural Computation*, vol. 1, pp. 541–551, 1989.
9. T. O'Shea, J. Corgan, and T. Clancy, "Convolutional Radio Modulation Recognition Networks," in *Proc. IEEE International Conference on Communications (ICC)*, 2016. Available: <https://arxiv.org/pdf/1602.04105>.
10. S. Ramabadran, A. S. Madhu Kumar, W. Guohua, and T. Shang Kee, "Blind Recognition of LDPC Code Parameters Over Erroneous Channel Conditions," *IET Signal Processing*, vol. 13, pp. 86–95, 2019, doi: 10.1049/iet-spr.2018.5025.
11. Zhang, X.; Zhang, W. A Cascade Network for Blind Recognition of LDPC Codes. *Electronics* 2023, 12, 1979. <https://doi.org/10.3390/electronics12091979>
12. Yan, W., Ling, Q., & Zhang, L. (2019). Convolutional Neural Networks for Space-Time Block Coding Recognition. *ArXiv, abs/1910.09952*.
13. Wang, T.; Yang, G.; Chen, P.; Xu, Z.; Jiang, M.; Ye, Q. A Survey of Applications of Deep Learning in Radio Signal Modulation Recognition. *Appl. Sci.* 2022, 12, 12052. <https://doi.org/10.3390/app122312052>
14. Wan, C.; Zhang, Q. A Novel Dual-Component Radar-Signal Modulation Recognition Method Based on CNN-ST. *Appl. Sci.* 2024, 14, 5499. <https://doi.org/10.3390/app14135499>
15. Y. Wang, J. Guo, H. Liu, L. Li, Z. Wang, and H. Wu, "CNN-Based Modulation Classification in the Complicated Communication Channel," in *Proc. IEEE Int. Conf. Electronic Measurement & Instruments (ICEMI)*, Yangzhou, China, Oct. 2017, pp. 512–516.
16. T. J. O'Shea, T. Roy and T. C. Clancy, "Over-the-Air Deep Learning Based Radio Signal Classification," in *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 1, pp. 168-179, Feb. 2018, doi: 10.1109/JSTSP.2018.2797022.
17. Rangaswamy, Arjun & T P, Surekha. (2021). Over-the-Air Modulation Classification using Deep Learning in Fading Channels for Cognitive Radio. *Indian Journal of Science and Technology.* 14. 3360-3369. 10.17485/IJST/v14i46.2073.
18. Yashashwi, Kumar & Sethi, Amit & Chaporkar, Prasanna. (2018). A Learnable Distortion Correction Module for Modulation Recognition. *IEEE Wireless Communications Letters.* 8. 10.1109/LWC.2018.2855749.
19. Erpek, T., O'Shea, T., Sagduyu, Y.E., Shi, Y., & Clancy, T.C. (2019). Deep Learning for Wireless Communications. *ArXiv, abs/2005.06068*.
20. O'Shea, Tim and Jakob Hoydis. "An Introduction to Deep Learning for the Physical Layer." *IEEE Transactions on Cognitive Communications and Networking* 3 (2017): 563-575.
21. Gu, Hao & Wang, Yu & Hong, Sheng & Gui, Guan. (2019). Blind Channel Identification Aided Generalized Automatic Modulation Recognition Based on Deep Learning. *IEEE Access.* PP. 1-1. 10.1109/ACCESS.2019.2934354.
22. S. Rajendran, W. Meert, D. Giustiniano, V. Lenders, and S. Pollin, "Distributed Deep Learning Models for Wireless Signal Classification with Low-Cost Spectrum Sensors," *IEEE Transactions on Cognitive Communications and Networking*, 2017, doi: 10.1109/TCCN.2018.2835460.

23. Q. Zhang, Z. Xu, and P. Zhang, "Modulation Recognition Using Wavelet-Assisted Convolutional Neural Network," in *Proc. IEEE Int. Conf. Advanced Technologies for Communications (ATC)*, Ho Chi Minh City, Vietnam, Oct. 2018, pp. 100–104.
24. F. Zhang, C. Luo, J. Xu, and Y. Luo, "An Efficient Deep Learning Model for Automatic Modulation Recognition Based on Parameter Estimation and Transformation," *IEEE Communications Letters*, 2021, doi: 10.1109/LCOMM.2021.3102656.
25. Peng, Shengliang & Jiang, Hanyu & Wang, Huaxia & Alwageed, Hathal & Yao, Yu-Dong. (2017). Modulation classification using convolutional Neural Network based deep learning model. 1-5. 10.1109/WOCC.2017.7929000.
26. N. E. West and T. O'Shea, "Deep Architectures for Modulation Recognition," in *Proc. IEEE DySPAN*, Baltimore, MD, USA, Mar. 2017, pp. 1–6.
27. R. Du, F. Liu, J. Xu, F. Gao, Z. Hu, and A. Zhang, "D-GF-CNN Algorithm for Modulation Recognition," *Wireless Personal Communications*, vol. 124, pp. 989–1010, 2022.
28. M. Zhang, M. Diao, and L. Guo, "Convolutional Neural Networks for Automatic Cognitive Radio Waveform Recognition," *IEEE Access*, vol. 5, pp. 11074–11082, 2017.
29. R. Li, L. Li, S. Yang, and S. Li, "Robust Automated VHF Modulation Recognition Based on Deep Convolutional Neural Networks," *IEEE Communications Letters*, vol. 22, pp. 946–949, 2018.
30. H. Wu, Q. Wang, L. Zhou, and J. Meng, "VHF Radio Signal Modulation Classification Based on Convolution Neural Networks," in *Proc. 1st Int. Symp. on Water System Operations, MATEC Web Conf.*, Beijing, China, Oct. 2018, vol. 246, p. 03032.
31. Y. Wang, M. Liu, J. Yang, and G. Gui, "Data-Driven Deep Learning for Automatic Modulation Recognition in Cognitive Radios," *IEEE Transactions on Vehicular Technology*, vol. 68, pp. 4074–4077, 2019.
32. M. Kulin, T. Kazaz, I. Moerman, and E. De Poorter, "End-to-End Learning from Spectrum Data: A Deep Learning Approach for Wireless Signal Identification in Spectrum Monitoring Applications," *IEEE Access*, vol. 6, pp. 18484–18501, 2018.
33. S. M. Hiremath, S. Deshmukh, R. Rakesh, and S. K. Patra, "Blind Identification of Radio Access Techniques Based on Time-Frequency Analysis and Convolutional Neural Network," in *Proc. IEEE TENCON*, Jeju Island, Korea, Oct. 2018, pp. 1163–1167.
34. H. S. Ghanem, R. M. Al-Makhlaway, W. El-Shafai, M. Elsabrouty, H. F. Hamed, G. M. Salama, and F. E. A. El-Samie, "Wireless Modulation Classification Based on Radon Transform and Convolutional Neural Networks," *Journal of Ambient Intelligence and Humanized Computing*, 2022.
35. P. Dileep, D. Das, and P. K. Bora, "Dense Layer Dropout Based CNN Architecture for Automatic Modulation Classification," in *Proc. IEEE NCC*, Kharagpur, India, Feb. 2020, pp. 1–5.
36. Z. Wang, D. Sun, K. Gong, W. Wang, and P. Sun, "A Lightweight CNN Architecture for Automatic Modulation Classification," *Electronics*, vol. 10, p. 2679, 2021.
37. K. Longi, T. Pulkkinen, and A. Klami, "Semi-Supervised Convolutional Neural Networks for Identifying Wi-Fi Interference Sources," in *Proc. Asian Conf. Machine Learning (ACML)*, Seoul, Korea, Nov. 2017, pp. 391–406.
38. Shi, Jibo, Qi, Lin, Li, Kuixian, Lin, Yun, Signal Modulation Recognition Method Based on Differential Privacy Federated Learning, *Wireless Communications and Mobile Computing*, 2021, 2537546, 13 pages, 2021. <https://doi.org/10.1155/2021/2537546>
39. S. Bhardwaj, D.-H. Kim, and D.-S. Kim, "Federated learning based modulation classification for multipath channels," *Parallel Computing*, vol. 120, p. 103083, 2024. doi:10.1016/j.parco.2024.103083.
40. DeepSig Inc. "Datasets." DeepSig, <https://www.deepsig.ai/datasets/>. Accessed 28 Mar. 2026.
41. M. Abudeeb, "RML2016.10b Dataset," *Kaggle*, <https://www.kaggle.com/datasets/marwanabudeeb/rml201610b>, accessed Mar. 28, 2026.
42. Tekbıyık, Kürşat, et al. "HisarMod: A New Challenging Modulated Signals Dataset." *IEEE DataPort*, 2020. Available: <https://iee-dataport.org/open-access/hisar-mod-new-challenging-modulated-signals-dataset>. Accessed: Mar. 28, 2026.
43. S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
44. D. Hong, Z. Zhang, and X. Xu, "Automatic Modulation Classification Using Recurrent Neural Networks," in *Proc. IEEE ICC*, Chengdu, China, Dec. 2017, pp. 695–700.

45. N. Daldal, Ö. Yıldırım, and K. Polat, "Deep Long Short-Term Memory Networks-Based Automatic Recognition of Six Different Digital Modulation Types Under Varying Noise Conditions," *Neural Computing and Applications*, vol. 31, pp. 1967–1981, 2019.
46. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention Is All You Need," in *Proc. NeurIPS*, 2017, pp. 6000–6010.
47. J. Cai, F. Gan, X. Cao *et al.*, "Signal Modulation Classification Based on the Transformer Network," *IEEE Transactions on Cognitive Communications and Networking*, vol. 8, no. 3, pp. 1348–1357, 2022.
48. N. Rashvand, K. Witham, G. Maldonado, V. Katariya, N. Marer Prabhu, G. Schirner, and H. Tabkhi, "Enhancing Automatic Modulation Recognition for IoT Applications Using Transformers," *IoT*, vol. 5, pp. 212–226, 2024, doi: 10.3390/iot5020011.
49. S. Lin, Y. Zeng, and Y. Gong, "Learning of Time-Frequency Attention Mechanism for Automatic Modulation Recognition," *IEEE Wireless Communications Letters*, vol. 11, pp. 707–711, 2022.
50. A. Dosovitskiy *et al.*, "An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale," arXiv:2010.11929, 2020.
51. Devlin, Jacob & Chang, Ming-Wei & Lee, Kenton & Toutanova, Kristina. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. 10.48550/arXiv.1810.04805.
52. Yenduri, G., Ramalingam, M., Selvi, G. C., Supriya, Y., Srivastava, G., Maddikunta, P. K. R., ... & Gadekallu, T. R. (2024). GPT (generative pre-trained transformer)—a comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions. IEEE access, 12, 54608-54649.
53. Touvron, Hugo, et al. "Llama: Open and efficient foundation language models." arXiv preprint arXiv:2302.13971 (2023).
54. Wei W, Zhu C, Hu L, Liu P. Application of a Transfer Learning Model Combining CNN and Self-Attention Mechanism in Wireless Signal Recognition. *Sensors (Basel)*. 2025 Jul 5;25(13):4202. doi: 10.3390/s25134202. PMID: 40648456; PMCID: PMC12252502.
55. Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. 2018. CBAM: Convolutional Block Attention Module. In *Computer Vision – ECCV 2018: 15th European Conference, Munich, Germany, September 8–14, 2018, Proceedings, Part VII*. Springer-Verlag, Berlin, Heidelberg, 3–19.
56. Thakur, A. S.; Imtiaz, M. Binary Transformer Detectors for Automatic Modulation Detection Under Realistic Radio Frequency Impairments. Preprints 2026, 2026030346. <https://doi.org/10.20944/preprints202603.0346.v1>
57. H. Khánh, S. Doan, and V.-P. Hoang, "Ensemble of Convolution Neural Networks for Improving Automatic Modulation Classification Performance," *Journal of Science and Technology – ICT Issue*, pp. 25–32, 2022, doi: 10.31130/ud-jst.2022.293E.
58. S. Zheng, P. Qi, S. Chen, and X. Yang, "Fusion Methods for CNN-Based Automatic Modulation Classification," *IEEE Access*, vol. 7, pp. 66496–66504, 2019.
59. L. Gao, X. Zhang, J. Gao, and S. You, "Fusion Image Based Radar Signal Feature Extraction and Modulation Recognition," *IEEE Access*, vol. 7, pp. 13135–13148, 2019.
60. H. Wu, Y. Li, L. Zhou, and J. Meng, "Convolutional Neural Network and Multi-Feature Fusion for Automatic Modulation Classification," *Electronics Letters*, vol. 55, pp. 895–897, 2019.
61. F. Shi, Z. Hu, C. Yue, and Z. Shen, "Combining Neural Networks for Modulation Recognition," *Digital Signal Processing*, vol. 120, p. 103264, 2022.
62. Ji, Zelin & Wang, Shuo & Yang, Kuojun & Zhang, Qinchuan & Ye, Peng. (2024). Transfer Learning Guided Noise Reduction for Automatic Modulation Classification. 10.48550/arXiv.2411.08376.
63. W. T. Zhang, D. Cui, and S. T. Lou, "Training Images Generation for CNN Based Automatic Modulation Classification," *IEEE Access*, vol. 9, pp. 62916–62925, 2021.
64. M. Zhang, Y. Zeng, Z. Han, and Y. Gong, "Automatic modulation recognition using deep learning architectures," in *Proc. IEEE 19th Int. Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Kalamata, Greece, Jun. 2018, pp. 1–5.
65. Vaidyanathan, P.P. (1990) Multirate Digital Filters, Filter Banks, Polyphase Networks, and Applications. A Tutorial. *Proceedings of the IEEE*, 78, 56-93. <https://doi.org/10.1109/5.52200>
66. Y. Sang and L. Li, "Application of novel architectures for modulation recognition," in *Proc. IEEE Asia Pacific Conf. on Circuits and Systems (APCCAS)*, Chengdu, China, Oct. 2018, pp. 159–162.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s)

disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.