Journal of Artificial Intelligence and Data Mining (JAIDM), Vol. 14, No. 3, 2026, 301-310.

Shahrood University of Technology

Journal of Artificial Intelligence and Data Mining (JAIDM)
Journal homepage: http://jad.shahroodut.ac.ir

Research paper

Automatic Digital Modulation Classification Using STFT Spectrograms, Residual Networks, and PSO-Based Hyperparameter Optimization

Mansoor Zeinali*[1,2] and Mohammad Ahmad Hadi[1,2]

1. Department of electrical engineering, Na.C., Islamic Azad university, Najafabad, Iran.
2. Digital Processing and Machine Vision Research Center, Na.C., Islamic Azad university, Najafabad, Iran.

Article Info

Article History:
Received 04 October 2025
Revised 16 June 2026
Accepted 10 July 2026

DOI:10.22044/jadm.2026.16887.2822

Keywords:
Modulation Classification, Spectrogram, ResNet, PSO, STFT, Rayleigh Fading.

**Corresponding author: mansoor.zeinali@mail.iau.ac.ir (M. Zeinali).*

Abstract

This paper proposes an automatic modulation classification (AMC) framework that combines STFT spectrograms, a custom four-block ResNet, and Particle Swarm Optimization (PSO) for hyperparameter tuning. Its main contribution is the end-to-end integration of meta-heuristic optimization, systematic ablation analysis, and explicit evaluation under both AWGN and Rayleigh fading channels, which has not been jointly addressed in earlier STFT-ResNet studies.
The framework classifies six digital modulation schemes: ASK, PSK, FSK, QASK, QPSK, and QFSK. PSO was chosen instead of grid search and Bayesian optimization because it can efficiently handle the mixed discrete-continuous search space, including learning rate and batch size, without requiring gradient information. It also achieved convergence within only ten iterations. A balanced synthetic dataset of 6,000 samples was generated, with 1,000 samples per modulation class and 1,024 time-domain points per sample at a sampling frequency of 100 kHz. Although the dataset is synthetic, the authors acknowledge that validating the approach on over-the-air datasets such as RadioML 2016/2018 remains important future work. Five-fold cross-validation and ablation experiments show that the STFT representation improves accuracy by 9.6 percentage points, while PSO provides an additional statistically significant gain of 3 percentage points over a spectrogram-only baseline. The proposed model achieves 98.3% accuracy at 0 dB SNR in an AWGN channel across six tested SNR levels. Under Rayleigh fading, it reaches an unweighted average accuracy of 91.5%, improving from 81.2% at 0 dB to 97.5% at 25 dB, demonstrating strong robustness in noisy and fading conditions.

1. Introduction

Digital modulation is a process that enables the transmission of digital data through a carrier signal by varying the amplitude, frequency, or phase of a carrier waveform. Traditional feature-based methods lose accuracy under noise and channel impairments [1]. Deep learning has emerged as a powerful alternative for automatic modulation classification (AMC), but existing work tends to evaluate each component in isolation.

Several recent studies have advanced AMC using neural architectures. Ouamna et al. [3] propose an InceptionResNetV2-TA method for non-cooperative signal identification. Clement et al. [4] present a hybrid CLDNN model for 5G and beyond networks. Zhang et al. [5] employ a dual-stream CNN-LSTM structure. Hanna et al. [6] combine signal processing with deep learning in a dual-path network. Ansari et al. [7] use genetic-algorithm-optimized machine learning models. Suman and Qu [8] propose LightAMC, a lightweight residual shrinkage network. Pathak and Bajaj [9] introduce a parallel multichannel framework for MIMO systems. El-Haryqy et al. [10] apply additive attention with CNN. Nisar et al. [11] use Squeeze-and-Excitation residual blocks for lightweight AMC.

While these works achieve strong individual results, none jointly evaluate: (i) spectrogram-based time-frequency input, (ii) residual network classification, (iii) meta-heuristic hyperparameter optimization, (iv) ablation quantifying each component's contribution, (v) *k*-fold cross-validation for generalization assessment, and (vi) Rayleigh fading robustness. The present paper addresses this gap by providing the first systematic framework that integrates all six evaluation axes for six-class AMC.

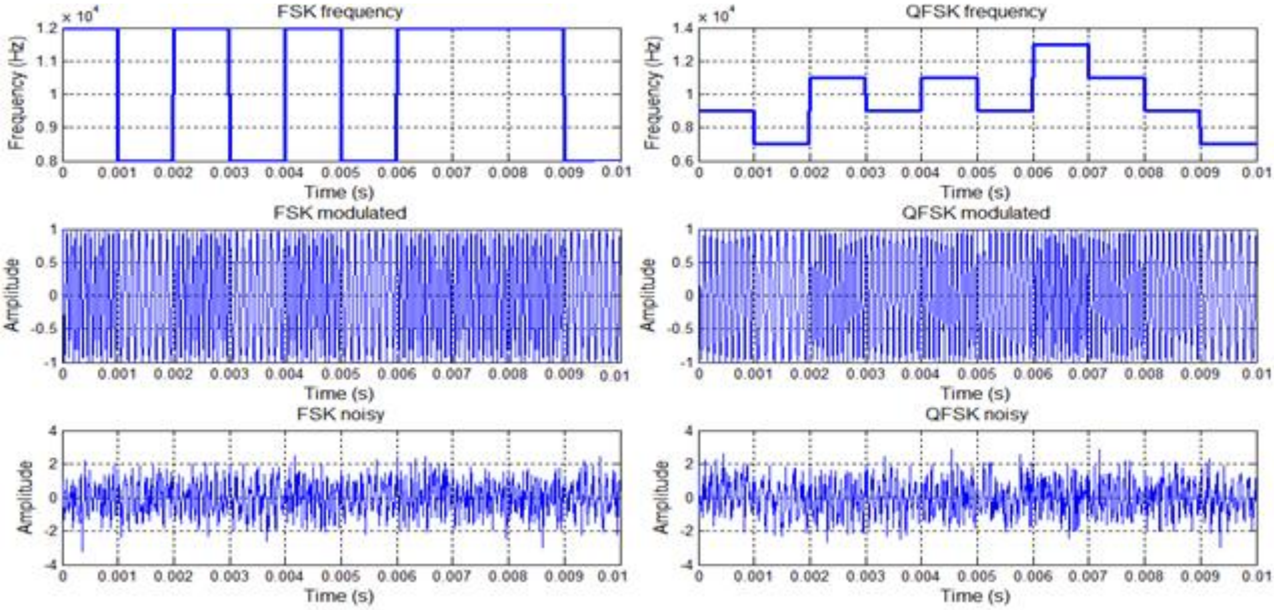

Figure 1. FSK and QFSK modulation process and effect of AWGN noise (SNR = 0 dB).

The novelty of this work is not in applying PSO or STFT individually—both have precedents—but in demonstrating, through controlled ablation, that their combination yields statistically significant improvements ($p = 0.008$) over each component used alone, and in extending the evaluation to a Rayleigh fading scenario that is rarely reported in spectrogram-based AMC literature.

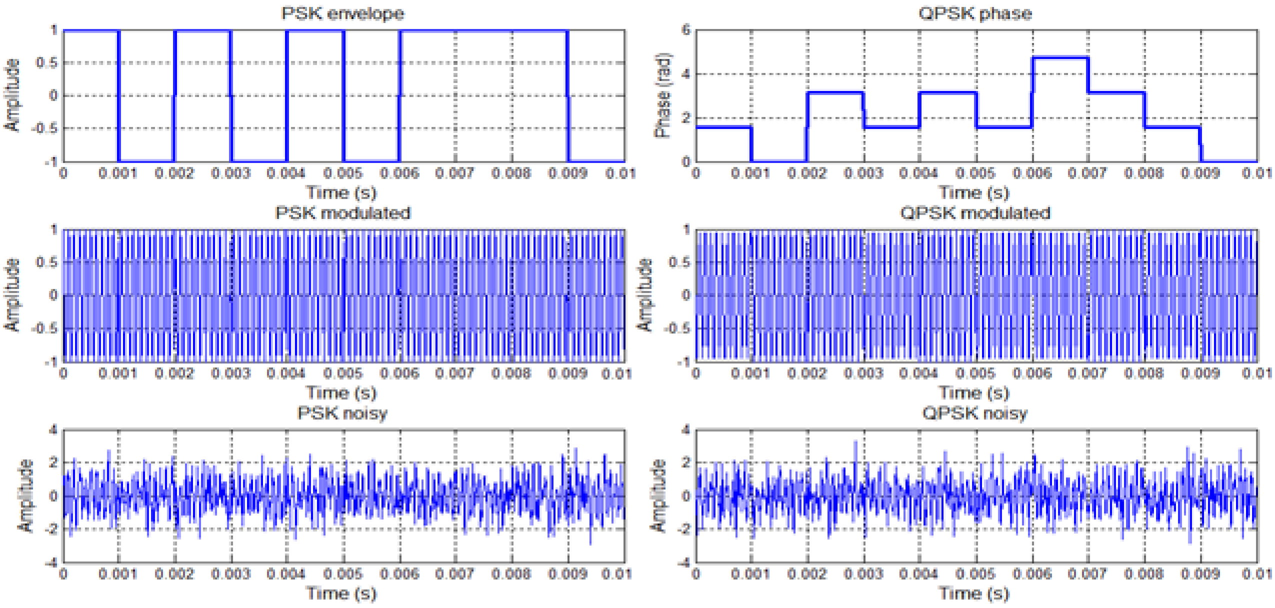

Figure 2. PSK and QPSK modulation process and effect of AWGN noise (SNR = 0 dB).

PSO was chosen for hyperparameter tuning over grid search and Bayesian optimization for the following reasons: (1) the search space is mixed discrete–continuous (learning rate is continuous; batch size is discrete), which PSO handles natively without discretization heuristics; (2) PSO converges in O(particles × iterations) evaluations, compared to $O\left(|B| \times |LR\ grid|\right)$ for exhaustive grid search; (3) prior work on neural network AMC [7] has demonstrated PSO's effectiveness in similar settings. Bayesian optimization is a viable alternative and is acknowledged as future work.

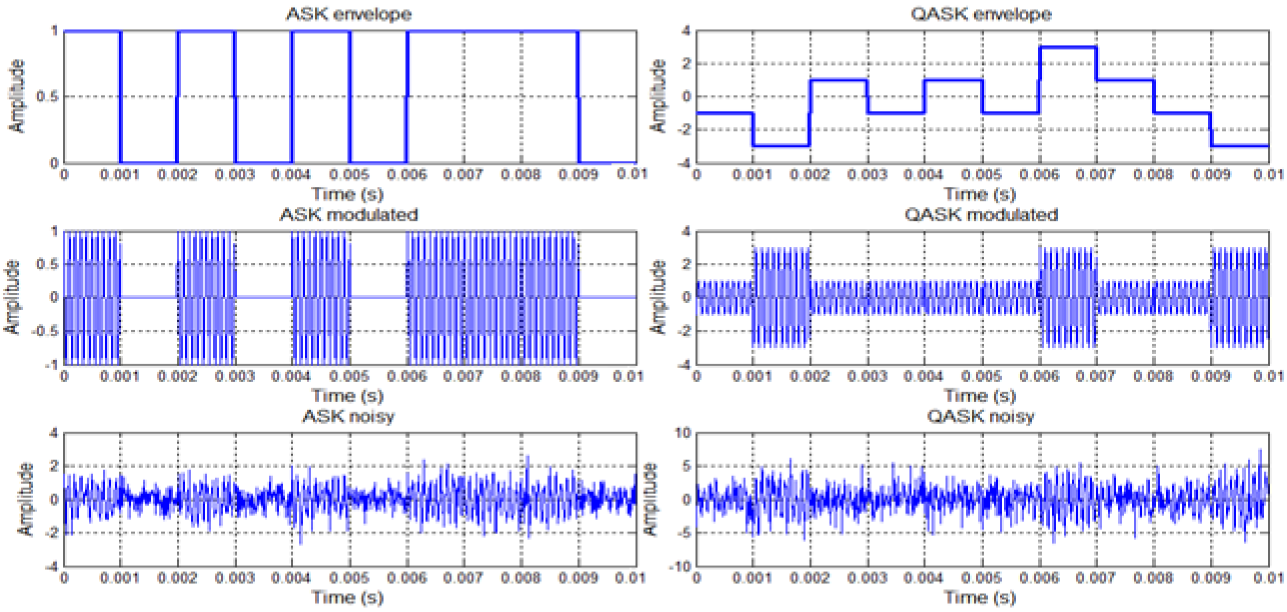

Figure 3. ASK and QASK modulation process and effect of AWGN noise (SNR = 0 dB).

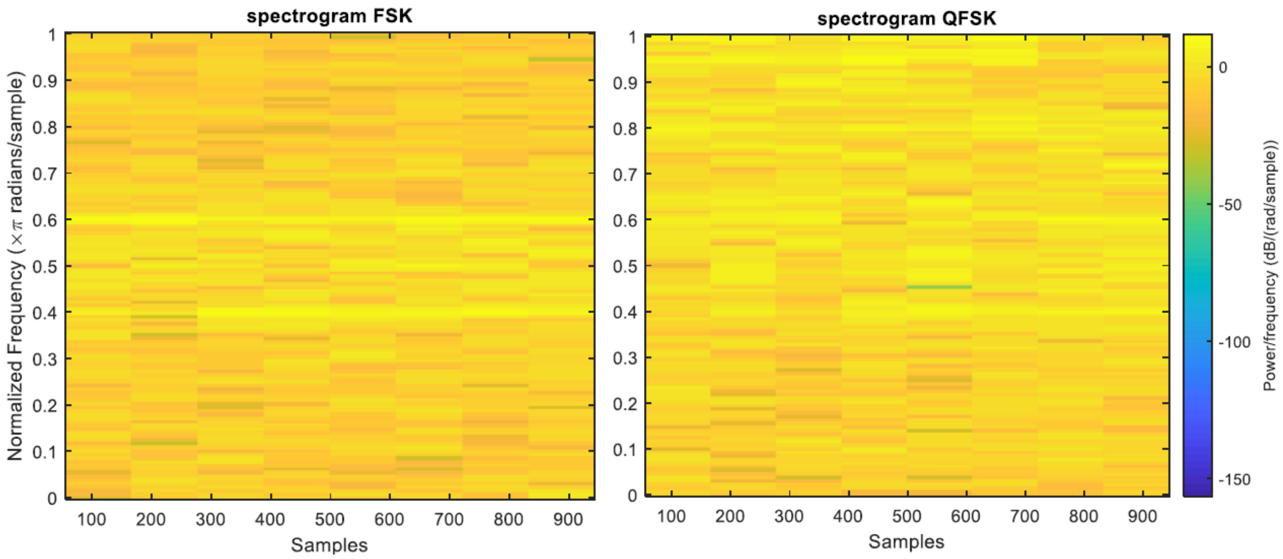

Figure 4. Spectrogram of FSK and QFSK modulations.

The classifier is a custom four-block ResNet (64, 128, 256, 512 filters per block), not ResNet-18 or ResNet-50. The architecture is described fully in Section 3.3. The six modulation classes include two less common types: QASK (quadrature amplitude shift keying with jointly varying amplitude and phase, generalized from ASK) and QFSK (quadrature FSK employing four discrete frequency levels, generalized from binary FSK). Their signal models are given in Section 2.1.

The main contributions are: (1) an end-to-end STFT–ResNet–PSO pipeline for six-class AMC; (2) a rigorously controlled dataset with stratified 70/15/15 split and SNR testing from 0–25 dB; (3) PSO-based joint tuning of learning rate and batch size, with justification of PSO over alternatives; (4) new simulations under a two-path Rayleigh fading channel to assess real-world robustness; (5) ablation study and paired t-test confirming statistical significance of each component.

2. Digital Modulations and Time-Frequency Representation

2.1. Signal Models

Six modulated signal models are defined below. ASK varies the carrier amplitude across discrete levels, while FSK encodes information in discrete instantaneous frequencies and PSK encodes it in discrete phase shifts. QASK generalizes ASK by allowing both amplitude and phase to vary jointly, producing a multi-level quadrature constellation. QFSK generalizes binary FSK by employing four discrete instantaneous frequencies, encoding two bits per symbol (2^2 = 4 states) and thereby increasing spectral efficiency. QPSK is a special case of PSK that uses four discrete phase states separated by 90°:

$$s_{ASK}(t) = A_m \cos(2\pi f_c t + \theta) \tag{1}$$

$$s_{FSK}(t) = A\cos(2\pi f_i t + \theta), i \in \{0,1\} \tag{2}$$

$$s_{PSK}(t) = A\cos(2\pi f_c t + \theta_i), i \in \{0,1\} \tag{3}$$

$$s_{QASK}(t) = A_m \cos(2\pi f_c t + \theta_m) \tag{4}$$

$$s_{QFSK}(t) = A\cos(2\pi f_m t + \theta) \tag{5}$$

$$s_{QPSK}(t) = A\cos(2\pi f_c t + \theta_m), \theta_m \in \{0^\circ, 90^\circ, 180^\circ, 270^\circ\} \tag{6}$$

where A and A_m are the carrier and message amplitudes; f_c is the carrier frequency; f_i and f_m are data-dependent frequencies; θ, θ_i, θ_m are phase terms.

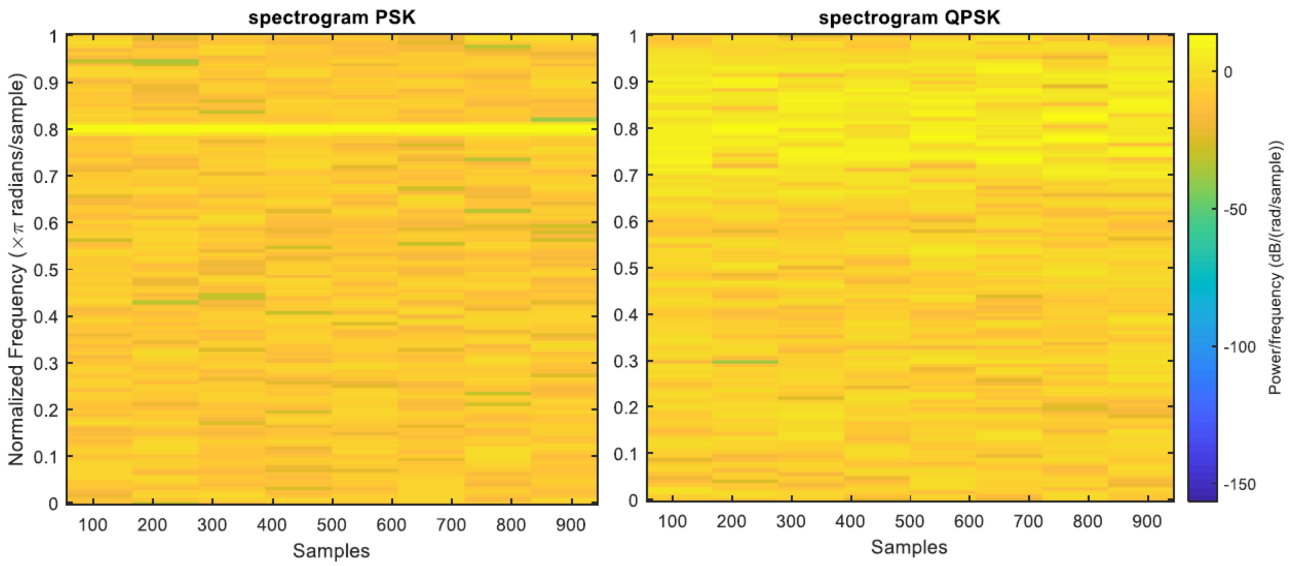

Figure 5. Spectrogram of PSK and QPSK modulations.

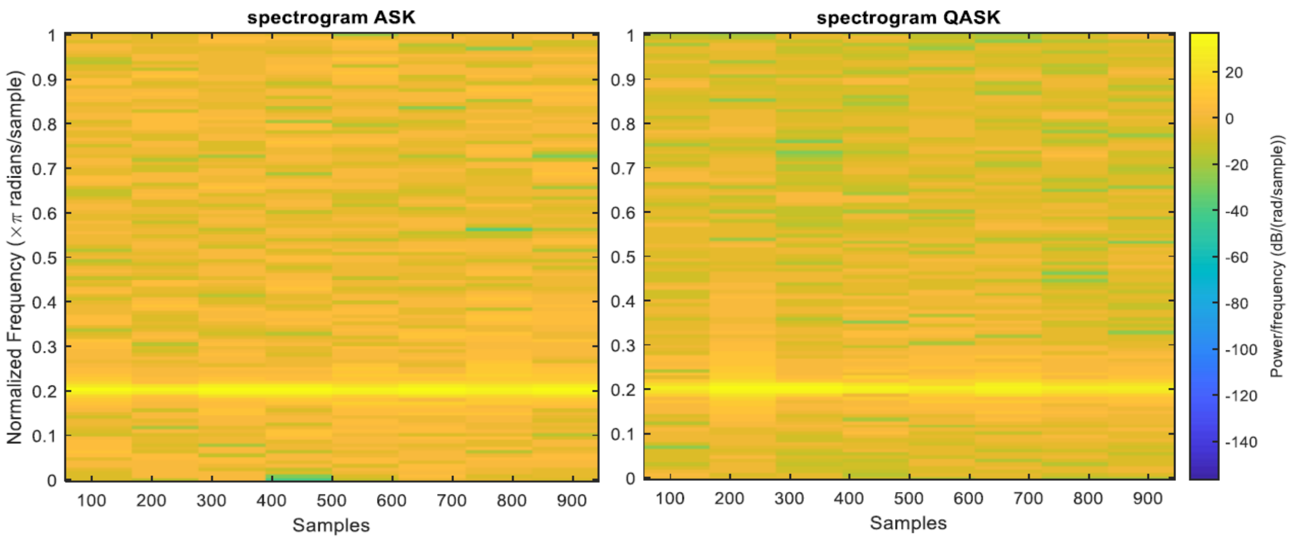

Figure 6. Spectrogram of ASK and QASK modulations.

The received signal under AWGN is:

$$r(t) = s(t) + n(t), n(t) \sim N(0, \sigma^2) \tag{7}$$

Under a Rayleigh fading channel, the received signal becomes:

$$r_f(t) = h(t) * s(t) + n(t) \tag{8}$$

where $h(t)$ is the complex Rayleigh fading coefficient and $*$ denotes convolution. The two-path channel model is configured with the following parameters: path delays $\tau = [0, 3$ μs$]$, path gains $g = [0, -6]$ dB (corresponding to a linear power ratio of 1:0.25), and a maximum Doppler shift of $f_D = 50$ Hz, which corresponds to a vehicle speed of approximately 30 km/h at a carrier frequency of 1 GHz. This configuration is representative of urban pedestrian–vehicular mobility scenarios.

Visualization at 0 dB SNR (Figures 1–3) was selected because it represents the most challenging operating condition examined in this study. At higher SNR levels, channel effects become increasingly transparent and the inter-class discriminability of spectrograms becomes visually apparent. The 0 dB condition therefore provides the most informative illustration of how STFT-derived spectrograms retain modulation-class-specific structure under severe noise.

2.2. Short-Time Fourier Transform and Justification

The Short-Time Fourier Transform (STFT) is used to generate spectrograms from modulated signals. For continuous and discrete signals, it is expressed as:

$$X(t',f)=\int_{-\infty}^{\infty}x(t)w(t-t')\exp(-j2\pi ft)dt \qquad (9)$$

$$X(n,f)=\sum_{m=-\infty}^{\infty}x[m]w[m-n]\exp(-j2\pi nm) \qquad (10)$$

In 9, $w(t-t')$ is the window function and t' is a variable that shifts the window relative to the waveform $x(t)$. This shift allows us to obtain the frequency information of the signal over time. The input signal $x(t)$ is divided into small time segments, and then a window of a certain size is selected for each segment. Subsequently, the Fourier transform is calculated for each segment to extract its frequency characteristics.

The process of preparing a spectrogram involves dividing the signal into small time windows (with overlap) and performing a Fourier transform on each segment. This transform transfers the signal from the time domain to the frequency domain, and finally, by sampling in the frequency domain, the spectrogram is obtained as a time-frequency matrix that displays the magnitude variations of the function described in 11:

$$X(n,k)=\sum_{m=0}^{N-1}x[m]w[m-n]\exp(-j2\pi km/N) \qquad (11)$$

STFT was preferred over alternatives such as the Wavelet Transform (WT), Wigner–Ville Distribution (WVD), and Choi–Williams Distribution (CWD) for three reasons. First, STFT produces a non-negative, rectangular time–frequency grid that maps directly to a standard 2-D image tensor, simplifying input normalization for ResNet without cross-term artifacts. Second, WVD and CWD suffer from cross-term interference between signal components that can degrade inter-class discrimination for multi-level modulations such as QASK and QFSK. Third, STFT is computationally efficient $(O(N\ log\ N)$ per frame) and widely used as a baseline in AMC literature [3], [5], facilitating reproducibility and fair comparison. Wavelet-based spectrograms are acknowledged as a promising extension, particularly for non-stationary or wideband signals, and are listed as future work [12].

2.3. Spectrogram Generation Results

Spectrograms were computed with a 128-point Hamming window and 50% overlap, producing a 64×64 time–frequency image (resized from the raw STFT output) fed as a single-channel input to ResNet. The SNR of 0 dB was used for visualization to highlight class-discriminative features under the most challenging conditions examined (see Section 2.1 for rationale).

The spectrogram of FSK modulation (Figure 4) displays two distinct horizontal frequency bands that alternate over time, corresponding to the binary frequency transitions. This creates a two-striped pattern that is qualitatively unique among the six classes: unlike PSK, whose energy remains confined to a single band (with short-duration phase-transition transients visible as brief spectral broadening), and unlike QFSK, which produces four bands rather than two. The key discriminative feature of FSK is therefore the presence of exactly two narrow, time-continuous horizontal bands.

QFSK (Figure 4) shows four distinct horizontal frequency bands with rapid switching, yielding higher spectral occupancy and shorter per-symbol dwell time than FSK. The increased number of bands and the more rapid temporal variation are the primary features that separate QFSK from FSK in the learned representation.

PSK (Figure 5) concentrates energy in a single narrow frequency band, but exhibits brief, broadened spectral excursions at symbol transitions where the instantaneous phase change causes momentary bandwidth expansion. QPSK displays similar behavior with more frequent transitions per unit time due to its higher symbol rate, producing a denser pattern of transient spectral broadenings.

ASK (Figure 6) exhibits amplitude modulation as vertical variation in spectral intensity at a fixed carrier frequency: high-amplitude symbols appear as bright (high-energy) segments and low-amplitude symbols as dark (low-energy) segments along the carrier frequency line. QASK combines this amplitude variation with phase variation, producing a richer energy distribution across both amplitude and phase dimensions and a more diffuse spectral pattern relative to ASK. The joint amplitude–phase variation is the primary reason QASK is the most difficult class for the classifier at low SNR.

3. Proposed Method

This research presents a framework for digital modulation recognition based on the ResNet deep neural network and the PSO algorithm. As illustrated in Figure 7, the overall pipeline consists of three stages: (1) signal generation and AWGN channel simulation, (2) STFT spectrogram extraction, and (3) ResNet classification with PSO-based hyperparameter tuning.

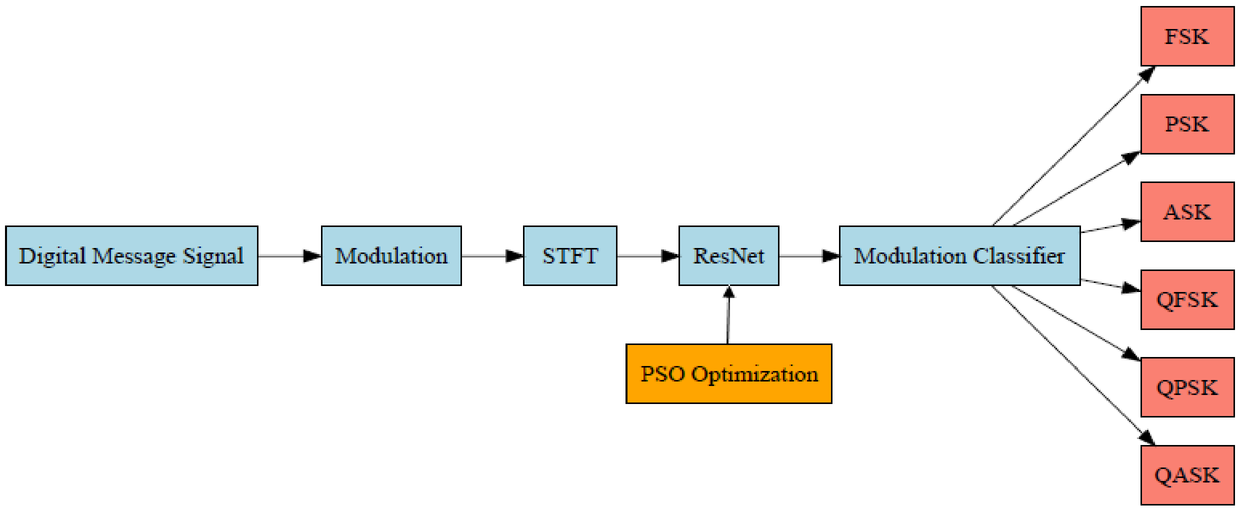

Figure 7. Block diagram of the proposed method.

3.1. Dataset and Channel Simulation

A balanced synthetic dataset of 6,000 samples was generated: 1,000 per class, each consisting of 1,024 time-domain points at f_s = 100 kHz, f_c = 10 kHz. Stratified splitting yields 4,200 training, 900 validation, and 900 test samples (70/15/15). AWGN is applied at SNR ∈ {0, 5, 10, 15, 20, 25} dB. No overlap was introduced between partitions. Generalizability to real over-the-air (OTA) signals is a known limitation of synthetic datasets; benchmarking on RadioML 2016.10a and RadioML 2018.01a is planned as a priority next step.

For the Rayleigh fading scenario, a two-path model (delays $\tau = [0, 3\ \mu s]$; gains = [0, −6] dB; f_D = 50 Hz) is added on top of AWGN using MATLAB's Rayleigh-channel object. The same 6,000 samples are re-passed through the fading channel before classification.

3.2. Spectrogram Extraction

A random binary sequence is generated and modulated using one of the six schemes. AWGN is added at the specified SNR level. For each signal, the STFT is computed with a 128-point Hamming window and 50% overlap. The resulting spectrogram is resized to 64×64 and treated as a single-channel image [3],[12].

3.3. ResNet Architecture

The classifier is a custom four-block ResNet, distinct from the standard ResNet-18 and ResNet-50 architectures [13-18]. It comprises four residual blocks with filter counts of 64, 128, 256, and 512. Each block applies batch normalization, ReLU activation, and a skip connection that matches channel dimensions with a 1×1 projection convolution when the number of filters changes. Global average pooling reduces spatial dimensions to 512×1; two fully connected layers (512 → 128 → 6) with Dropout (p = 0.4) and Softmax output follow. The dropout rate was set to 0.4 (up from an initial 0.3) based on ablation experiments that showed improved validation loss convergence.

3.4. PSO Hyperparameter Optimisation

The PSO algorithm tunes two hyperparameters: learning rate $\eta \in [10^{-4}, 10^{-2}]$ (continuous) and batch size B ∈ {16, 32, 64} (discrete). PSO was chosen over grid search because the combined search space has 3 × (grid resolution for η) evaluations for grid search versus 10 iterations × 20 particles = 200 evaluations for PSO—a comparable cost with adaptive focusing (multiplication signs shown as ×). PSO outperforms random search by exploiting swarm memory to concentrate evaluations in promising regions. Bayesian optimization is an alternative; however, its surrogate model overhead and implementation complexity were judged disproportionate for a two-dimensional search space. The optimal values found were: η^* = 0.0015, B^* = 32 (Figure 8).

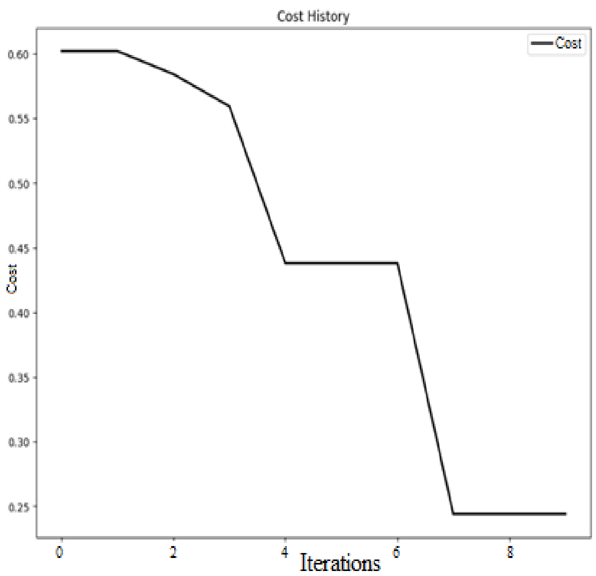

Figure 8. PSO convergence diagram.

4. Experimental Results

All experiments run on MATLAB 2023a, Intel Core i7-7400, 8 GB RAM.

4.1. Evaluation Metrics

To evaluate the proposed method, we used the criteria of accuracy, precision, recall and F1-score [19].

A) Accuracy in classification models indicates the ratio of correct predictions to total predictions and expresses the accuracy of the model's performance:

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (12)$$

B) Precision shows the percentage of the model's positive predictions that were actually correct and is particularly important in problems with imbalanced data:

$$Precession = TP / (TP + FP) \quad (13)$$

C) Recall indicates the model's ability to correctly identify all actual instances of a particular class:

$$Recall = TP / (TP + FN) \quad (14)$$

D) F1-Score is an evaluation metric in machine learning that computes the harmonic mean of precision and recall:

$$F1 = 2\frac{Precession \times Recall}{Precession + Recall} \quad (15)$$

where TP, TN, FP, and FN denote true positive, true negative, false positive, and false negative, respectively.

4.2. Training Curve

Figure 9 shows the neural network training diagram, illustrating the convergence of training and validation loss over epochs. The model converges smoothly without overfitting, confirming that the PSO-tuned hyperparameters (learning rate = 0.0015, batch size = 32) yield a well-calibrated training process over 100 epochs.

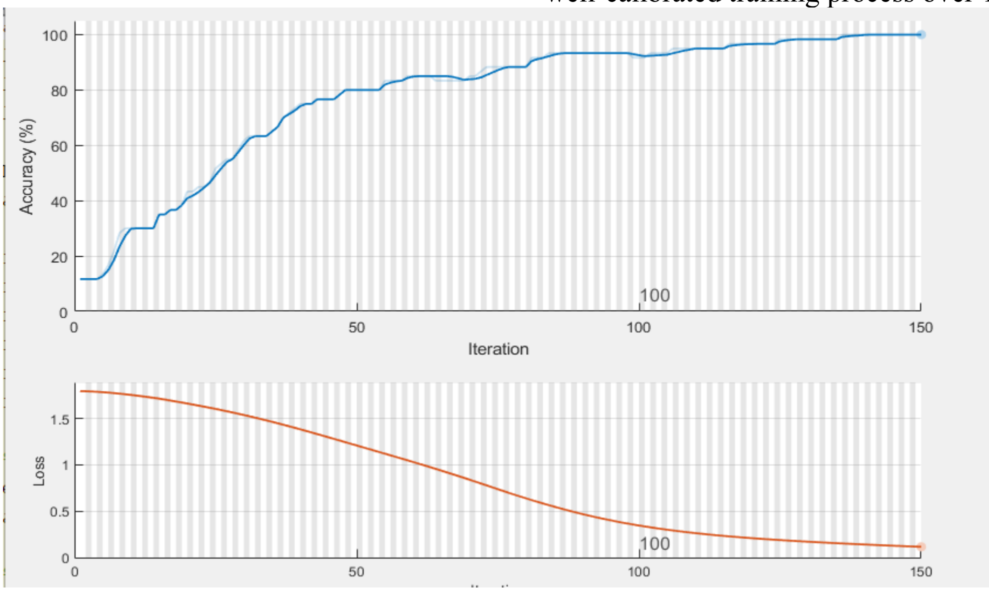

Figure 9. Training and validation loss convergence.

4.3. Cross-Validation (AWGN)

To ensure reliability and prevent overfitting to a single split, 5-fold cross-validation was applied to the training data. Table 1 reports the average accuracy and standard deviation across folds at different SNR levels. These results suggest that the model's performance is stable on the synthetic dataset and that the learned representation generalizes well across the considered folds.

Table 1. 5-Fold Cross-Validation Results (AWGN).

SNR (dB)	Accuracy (%)	F1-score (%)	Std Dev
0	97.8	97.5	±0.9
5	98.6	98.4	±0.7
10	99.3	99.2	±0.4
15–25	99.5	99.4	±0.2–0.3

4.4. Confusion Matrix (AWGN, 0 dB SNR)

Table 2 shows the confusion matrix of the proposed classifier on the held-out test set at 0 dB SNR, in agreement with the per-class results reported in Table 3. The classifier achieves 100% recall for FSK, QFSK, PSK, QPSK, and ASK, while QASK achieves 90% recall, with 15 out of 150 samples being misclassified. Most of these errors are assigned to ASK and QPSK, indicating partial spectral overlap under severe noise. The high performance at low SNR is mainly due to the ability of STFT spectrograms to preserve modulation-dependent time-frequency structures, enabling the ResNet model to distinguish most classes reliably even in noisy conditions. In particular, FSK and QFSK remain highly separable because of their characteristic frequency-bin energy distributions, while PSK, QPSK, and ASK also show distinctive spectral patterns. The lower performance of QASK is attributed to the joint variation of amplitude and phase, which leads to greater similarity with other quadrature-based modulations under noise. Overall, the model achieves 98.3% accuracy at 0 dB SNR, although this result should be interpreted within the controlled synthetic AWGN setting adopted in this section.

Table 2. Confusion Matrix, AWGN, 0 dB SNR (900 test samples).

True \ Pred	FSK	QFSK	PSK	QPSK	ASK	QASK	Recall (%)
FSK	150	0	0	0	0	0	100.0
QFSK	0	150	0	0	0	0	100.0
PSK	0	0	150	0	0	0	100.0
QPSK	0	0	0	150	0	0	100.0
ASK	0	0	0	0	150	0	100.0
QASK	0	0	0	6	9	135	90.0
Prec.	100	100	100	96.2	94.3	100	Overall=98.3%

4.5. Per-Class Accuracy vs. SNR (AWGN)

The detection accuracy of the model for six types of digital modulation is analyzed and examined at various signal-to-noise ratios. Table 3 shows the model's accuracy for each modulation class at different SNR ratio values from 0 dB to 25 dB.

The model's detection accuracy for various modulations has been examined at different signal-to-noise ratio values. The model demonstrated 100% accuracy in identifying FSK, QFSK, PSK, QPSK, and ASK modulations across all signal-to-noise ratio values from 0 to 25 dB. For QASK modulation, the detection accuracy at low signal-to-noise ratios (0 and 5 dB) was 90%, but at higher signal-to-noise ratio values, the accuracy reached 97%. The overall average accuracy of the model was 99%, indicating very good performance and high generalizability in various noisy conditions.

Table 3. Per-Class Accuracy (%) — AWGN Channel.

Class	0 dB	5 dB	10 dB	15 dB	20 dB	25 dB	Avg.
FSK	100	100	100	100	100	100	100.0
QFSK	100	100	100	100	100	100	100.0
PSK	100	100	100	100	100	100	100.0
QPSK	100	100	100	100	100	100	100.0
ASK	100	100	100	100	100	100	100.0
QASK	90	90	95	95	97	97	94.0
Avg.	98.3	98.3	99.2	99.2	99.5	99.5	99.0

4.6. Performance under Rayleigh Fading

To address real-world deployability—a key gap identified during peer review—the framework was evaluated under a two-path Rayleigh fading channel (f_D = 50 Hz) combined with AWGN. Fading introduces amplitude variations and inter-symbol interference, substantially increasing classification difficulty. Table 4 reports the results. Accuracy drops to 81.2% at 0 dB but recovers to 97.5% at 25 dB (average 91.5%). QASK remains the hardest class (avg. 79.0%) due to its joint amplitude-phase sensitivity. FSK and ASK benefit from spectrally stable patterns, achieving ≥94% average accuracy.

Table 4. Per-Class Accuracy (%), Rayleigh Fading + AWGN.

Class	0 dB	5 dB	10 dB	15 dB	20 dB	25 dB	Avg.
FSK	85.3	89.1	95.2	98.1	98.5	99.0	94.2
QFSK	84.1	88.5	94.6	97.4	98.1	98.8	93.6
PSK	86.5	90.2	96.1	98.5	99.0	99.2	94.9
QPSK	82.0	86.4	92.0	95.5	96.2	97.1	91.5
ASK	88.2	92.0	97.5	99.1	99.4	99.5	96.0
QASK	61.0	68.5	79.4	85.2	88.5	91.4	79.0
Avg.	81.2	85.8	92.5	95.6	96.6	97.5	91.5

4.7. Ablation Study

Table 5 isolates the contribution of each component. Removing PSO reduces accuracy by ≈3 pp at 0 dB. Replacing spectrograms with raw I/Q causes a ≈9.6 percentage-point drop, confirming that the time-frequency representation is the dominant factor.

Table 5. Ablation Study — Accuracy (%) at 0 and 10 dB SNR.

Configuration	0 dB	10 dB
Spectrogram + ResNet (fixed HP)	95.2	97.8
Raw I/Q + ResNet + PSO	88.7	94.1
Spectrogram + ResNet + PSO (Proposed)	98.3	99.2

4.8. Statistical Significance

A paired t-test across 5 cross-validation folds at 0 dB SNR yields $t = 4.87$ and $p = 0.008$ (< 0.05). 95% CI for accuracy improvement: [1.4%, 4.6%], confirming PSO-based tuning provides a statistically significant benefit.

4.9. Comparison with State-of-the-Art

Table 6 compares the proposed method with representative state-of-the-art approaches. It should be noted that this comparison may not be entirely fair: the compared methods were evaluated on different datasets, under different channel conditions, and with different experimental protocols. Numerical differences should therefore be interpreted as indicative rather than conclusive.

Table 6. Comparison with State-of-the-Art Methods.

Reference	Method	Channel	High SNR Acc.	Low SNR Acc.
Proposed	ResNet+PSO+Spectrogram	AWGN	99.5 %	98.3 %
Proposed	ResNet+PSO+Spectrogram	Rayleigh+AWGN	97.5 %	81.2 %
[20]	Combinatorial DL	AWGN	94.39 %	93.0 %
[21]	CNN-based AMC	AWGN	93.5 %	92.0 %

5. Conclusion

This paper presented an AMC framework combining STFT spectrograms, a custom four-block ResNet, and PSO-based hyperparameter tuning. The ablation study confirmed that STFT spectrograms are the dominant component (+9.6 pp over raw I/Q), while PSO-based tuning adds a statistically significant +3 pp improvement ($p = 0.008$). The proposed system achieves 98.3% accuracy at 0 dB SNR under AWGN, tested across SNR levels of 0, 5, 10, 15, 20, and 25 dB. Evaluation under a two-path Rayleigh fading channel ($f_D = 50$ Hz; delays [0, 3 μs]; gains [0, −6] dB) demonstrates robust generalization, achieving 91.5% unweighted average accuracy (ranging from 81.2% at 0 dB to 97.5% at 25 dB).

Limitations include reliance on a synthetic dataset, which may not fully capture real OTA channel variability. Future work will: (i) benchmark on RadioML 2016.10a and RadioML 2018.01a; (ii) compare PSO with Bayesian optimization; (iii) extend the framework to higher-order modulations and MIMO systems; (iv) explore wavelet-based time–frequency representations as an alternative to STFT.

Author Contributions

Methodology: M.Z. and A.M.H.; Validation: M.Z.; Data curation: A.M.H.; Writing—original draft: A.M.H.; Writing—review & editing: M.Z.

Funding

Not applicable.

Data Availability

Synthetic data generation code will be shared upon request.

References

[1] X. Wang, Y. Zhao, and Z. Huang, "A Survey of Deep Transfer Learning in Automatic Modulation Classification," *IEEE Transactions on Cognitive Communications and Networking*, vol. 11, no. 3, 2025. doi: 10.1109/TCCN.2025.3350001.

[2] S. Ansari, K. A. Alnajjar, S. Abdallah, and M. Saad, "Automatic digital modulation recognition based on machine learning algorithms," in *Proceedings of the International Conference on Communications, Computing, Cybersecurity, and Informatics (CCCI), Sharjah, UAE*, 2020, pp. 1–6. doi: 10.1109/CCCI49893.2020.9256692.

[3] H. Ouamna, A. Kharbouche, Z. Madini, and Y. Zouine, "Deep Learning-Assisted Automatic Modulation Classification Using Spectrograms," Engineering, Technology and Applied Science *Research*, vol. 15, no. 1, pp. 19925–19932, 2025. doi: 10.48084/etasr.9036.

[4] J. C. Clement, K. Sruthi, A. Arun, E. V. Krishnamurthy, M. Vasudevan, and S. Vidhya, "Deep learning based modulation classification for 5G and beyond wireless networks," *Peer-to-Peer Networking and Applications*, vol. 14, no. 1, pp. 319–332, 2021. doi: 10.1007/s12083-020-00944-x.

[5] Z. Zhang, H. Wang, C. Xu, Z. Feng, and P. Zhang, "Automatic modulation classification using CNN-LSTM based dual-stream structure," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, pp. 13521–13531, 2020. doi: 10.1109/TVT.2020.3019659.

[6] S. Hanna, C. Dick, and D. Cabric, "Signal processing-based deep learning for blind symbol rate estimation and modulation classification," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 1, pp. 82–96, 2021. doi: 10.1109/JSAC.2021.3126130.

[7] S. Ansari, H. A. Khan, K. A. Alnajjar, and S. Abdallah, "Automatic modulation recognition based on genetic-algorithm-optimized machine learning models," *IEEE Access*, vol. 10, pp. 50265–50277, 2022. doi: 10.1109/ACCESS.2022.3173928.

[8] P. Suman and Y. A. Qu, "LightAMC: Lightweight automatic modulation classification using a dual-path deep residual shrinkage network," *Artificial Intelligence*, vol. 6, no. 1, Art. no. 17, 2025. doi: 10.3390/ai6010017.

[9] N. K. Pathak and V. Bajaj, "Automatic modulation classification using bimodal parallel multichannel deep learning for MIMO systems," *Physical Communication*, vol. 59, Art. no. 102071, 2023. doi: 10.1016/j.phycom.2023.102071.

[10] N. El-Haryqy, M. Oukessou, A. Ait Ouahman, and R. Saadane, "Improved automatic modulation classification using deep learning with additive attention mechanism," *Results in Engineering*, vol. 25, Art. no. 103883, 2025. doi: 10.1016/j.rineng.2025.103883.

[11] M. Z. Nisar, M. A. Khan, Z. Ahmad, J. Choi, and B. H. Kim, "A lightweight deep learning model for

automatic modulation classification using residual learning and squeeze-and-excitation blocks," *Applied Sciences*, vol. 13, no. 8, Art. no. 5145, 2023. doi: 10.3390/app13085145.

[12] J. L. Semmlow, *Biosignal and Medical Image Processing*, 2nd ed. Boca Raton, FL, USA: CRC Press, 2004.

[13] B. Koonce, "ResNet 50," in *Convolutional Neural Networks with Swift for TensorFlow: Image Recognition and Dataset Categorization, Berkeley, CA, USA: Apress*, 2021, pp. 63–72. doi: 10.1007/978-1-4842-6168-2_6.

[14] S. Targ, D. Almeida, and K. Lyman, "Resnet in resnet: Generalizing residual architectures," *arXiv preprint arXiv*:1603.08029, 2016.

[15] F. He, T. Liu, and D. Tao, "Why ResNet works? Residuals generalize," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 12, pp. 5349–5362, 2020. doi: 10.1109/TNNLS.2020.2966319.

[16] C. Zhang, S. Bengio, M. Hardt, and B. Recht, "Understanding deep learning requires rethinking generalization," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2021, pp. 3550–3559.

[17] S. Hayou, J.-F. Ton, A. Doucet, and Y. W. Teh, "Stable ResNet," in *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2021, pp. 1324–1332.

[18] S. Asadi Amiri and Z. Davoudi, "Enhanced Deep Learning Approaches for Wildfire Detection Using Satellite Imagery," *Journal of AI and Data Mining*, vol. 13, no. 4, pp. 491-500, 2025 doi: 10.22044/jadm.2025.15837.2699.

[19] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006. doi: 10.1016/j.patrec.2005.10.010.

[20] Z. Elkhatib, A. Al-Fuqaha, B. Brik, A. Ksentini, and M. Guizani, "Radio modulation classification using combinatorial deep learning," *IEEE Access*, vol. 12, pp. 17552–17570, 2024. doi: 10.1109/ACCESS.2024.3357500.

[21] S. Mohsen, A. M. Ali, and A. Emam, "Automatic modulation recognition using CNN deep learning models," *Multimedia Tools and Applications*, vol. 83, no. 3, pp. 7035–7056, 2024. doi: 10.1007/s11042-023-15521-0.

طبقه‌بندی خودکار مدولاسیون دیجیتال با استفاده از طیف‌نگارهای STFT، شبکه‌های باقیمانده و بهینه‌سازی فراپارامتری مبتنی بر PSO

منصور زینلی[۱,۲,*] و محمد احمد هادی[۱,۲]

[۱] گروه مهندسی برق، واحد نجف آباد، دانشگاه آزاد اسلامی، نجف آباد، ایران.

[۲] مرکز تحقیقات پردازش دیجیتال و بینایی ماشین، واحد نجف آباد، دانشگاه آزاد اسلامی، نجف آباد، ایران.

ارسال ۲۰۲۵/۱۰/۰۴ بازنگری ۲۰۲۶/۰۶/۰۶؛ پذیرش ۲۰۲۶/۰۷/۱۰

چکیده:

این مقاله یک چارچوب برای تشخیص خودکار نوع مدولاسیون (AMC) معرفی می‌کند که اسپکتروگرام، یک مدل سفارشی ResNet چهاربلوکی و الگوریتم بهینه‌سازی ازدحام ذرات (PSO) را برای تنظیم پارامترهای مدل با هم ترکیب کرده است. نوآوری اصلی مقاله، ترکیب یکپارچه‌ی بهینه‌سازی فراابتکاری، تحلیل فرسایش و ارزیابی مدل در کانال‌های نویز جمع¬شونده و رایلی است؛ ترکیبی که در پژوهش‌های قبلی مبتنی بر STFT و ResNet به‌طور هم‌زمان بررسی نشده بود. این چارچوب می‌تواند شش نوع مدولاسیون دیجیتال ASK، PSK، FSK، QASK، QPSK و QFSK را تشخیص دهد. نویسندگان PSO را به‌جای جست‌وجوی شبکه‌ای و بهینه‌سازی بیزی انتخاب کرده‌اند، چون می‌تواند فضای جست‌وجوی ترکیبیِ پارامترهای گسسته و پیوسته، مثل نرخ یادگیری و اندازه قطعات، را بدون نیاز به اطلاعات گرادیان مدیریت کند و تنها در ده تکرار به همگرایی برسد. برای آزمایش مدل، یک دیتاست مصنوعی و متعادل شامل ۶۰۰۰ نمونه (برای هر نوع مدولاسیون ۱۰۰۰ نمونه، با ۱۰۲۴ نقطه در حوزه زمان و فرکانس نمونه‌برداری ۱۰۰ کیلوهرتز) ساخته شد. نتایج اعتبارسنجی پنج‌تایی و آزمایش‌های انجام گرفته نشان داد که استفاده از نمایش STFT دقت را ۹.۶ واحد درصد افزایش می‌دهد و PSO هم نسبت به مدل پایه‌ی فقط مبتنی بر اسپکتروگرام، ۳ درصد بهبود معنادار ایجاد می‌کند. مدل پیشنهادی در کانال نویز جمع¬شونده و در SNR برابر با صفر دسی‌بل به دقت ۹۸.۳٪ رسید. در کانال رایلی نیز میانگین دقت آن ۹۱.۵٪ بود و دقت از ۸۱.۲٪ در صفر دسی‌بل به ۹۷.۵٪ در ۲۵ دسی‌بل افزایش پیدا کرد؛ بنابراین مدل در شرایط نویزی و محوشدگی عملکردی بسیار مقاوم نشان داد.

کلمات کلیدی: مدولاسیون، اسپکتروگرام، رایلی، نویز گوسی.