Hybrid CNN-BiLSTM Network with SNR-Aware Fusion for Automatic Modulation Classification

Sihem SOUIKI [1,6], Mohammed Hicham HACHEMI [2,6], Mohammed MERZOUG [3,7], Mourad HADJILA [4,6], Mohammed M'HAMEDI [5,7], Lamia BENLALDJ [4,6]

[1]Dept. of Electronic and Telecommunication, University of Ain-Temouchent, Ain-Temouchent 46000, Algeria
[2] Dept. of Electronics, Faculty of Electrical Engineering, University of Science and Technology of Oran-Mohamed Boudiaf, Oran 31000, Algeria
[3] Dept. of Computer Science, Faculty of Science, University Abou-bakr Belkaid, Tlemcen 13000, Algeria
[4] Dept. of Telecommunication, University Abou-bakr Belkaid, Tlemcen 13000, Algeria
[5]Dept. of Second Cycle, Ecole Supérieure en Sciences Appliquées de Tlemcen ESSAT, Tlemcen 13000, Algeria
[6]STIC Laboratory, Faculty of Technology, University Abou-Bakr -Belkaid, Tlemcen 13000, Algeria
[7]LRIT Laboratory, Faculty of Science, University Abou-bakr Belkaid, Tlemcen 13000, Algeria

sihem.souiki@univ-temouchent.edu.dz, hicham.hachemi@univ-usto.dz,
{mohammed.merzoug, mourad.hadjila,m.mhamedi, lamia.benlaldj}@univ-tlemcen.dz,

Submitted July 16, 2026 / Accepted October 10, 2026 / Online first November 20, 2026

Abstract. *Automatic Modulation Recognition (AMR) is an important component in cognitive radio and spectrum monitoring, where robust classification under varying signal-to-noise ratio (SNR) conditions remains challenging. This paper proposes a compact hybrid deep learning architecture that combines spatial-temporal feature extraction with explicit channel-condition awareness. The proposed network processes raw In-phase and quadrature (I/Q) samples using a hierarchical 1D convolutional neural network (CNN) to extract local features, followed by a bidirectional long short-term memory (BiLSTM) block to capture temporal dependencies. Concurrently an independent SNR-processing branch transforms the scalar SNR value into a low-dimensional embedding and concatenates it with the learned CNN–BiLSTM representation before classification. The model is evaluated on RadioML2016.10a and RadioML2016.10b across SNR levels from –20 dB to +18 dB. It achieves overall accuracies of 67.23% and 68.52%, average low–SNR accuracies of 26.20% and 28.93% in the –20 to –8 dB range, and a compact profile of 4.63 MFLOPs and 0.804M parameters. These results indicate improved AMC performance under controlled synthetic benchmark conditions, while real RF validation remains necessary before practical deployment.*

Keywords

Automatic modulation classification, deep learning, CNN–BiLSTM hybrid model, SNR-aware fusion, I/Q signal processing, RadioML datasets

1. Introduction

Automatic Modulation Classification (AMC) has become a fundamental component of intelligent wireless communication systems due to its critical role in spectrum monitoring, cognitive radio, electronic warfare, and adaptive communication networks [1], [2]. With the rapid evolution of modern wireless technologies, including the Internet of Things (IoT), fifth-generation (5G), and emerging sixth-generation (6G) communication systems, wireless environments are becoming increasingly dense, heterogeneous, and dynamically changing [3]. In such scenarios, the capability to automatically recognize the modulation format of received signals without prior knowledge of transmission parameters is essential for efficient spectrum utilization, interference mitigation, and reliable communication management [2], [4].

Despite considerable advances in AMC research, achieving robust modulation classification under practical wireless conditions remains a challenging task, particularly when models must generalize beyond controlled benchmark datasets to real over-the-air environments. Wireless signals are frequently affected by noise, fading, interference, and channel distortions, especially in low–SNR environments where identifying discriminative signal characteristics become challenging. Traditional methods that rely on creating specific features need manual design and expert knowledge, which makes it challenging for them to adapt and perform well in different channel conditions [5], [6]. Although deep learning (DL) approaches have recently shown better results by allowing automatic feature extraction directly from raw in-phase and quadrature (*I/Q*) samples. However, many current architectures still have trouble capturing local spatial features, long-range temporal dependencies, and channel condition awareness all in one framework [7–10].

DOI: 10.13164/re.2026.0543

Recent studies have shown that convolutional neural networks (CNNs) are effective for extracting discriminative spatial features from *I/Q* signal representations, while recurrent architectures such as long short-term memory (LSTM) and bidirectional LSTM (BiLSTM) networks can efficiently model temporal dependencies inherent in communication signals [11], [12]. More recently, attention mechanisms and Transformer–based architectures have emerged as promising solutions for AMC by enabling the modeling of both local and global contextual dependencies within wireless signal sequences [13]. Several recent works have explored hybrid CNN–Transformer and attention-enhanced frameworks to improve feature extraction capability, robustness under low-SNR conditions, and computational efficiency for real-time deployment. Nevertheless, many existing approaches still suffer from increased computational complexity or reduced generalization capability in severely noisy benchmark conditions, which may limit their potential applicability in resource-constrained wireless communication systems [14–19]. Although CNN–, LSTM–, and CNN–BiLSTM–based architectures have been widely investigated for AMC, these architectures alone do not fully address the need for explicit channel-condition awareness. Most existing models rely mainly on raw *I/Q* samples and attempt to learn noise-dependent behavior implicitly from the waveform. In contrast, the proposed framework introduces an independent SNR-processing branch that explicitly encodes channel-quality information and fuses it with the learned CNN–BiLSTM spatial-temporal representation at the feature level. Therefore, the novelty of this work does not lie in the use of CNN or BiLSTM layers alone, but in the compact dual-branch CNN–BiLSTM–SNR design and its explicit feature-level fusion between waveform features and SNR-based channel-condition information.

To address these limitations, this paper proposes a hybrid DL architecture that combines CNNs, BiLSTM networks, and an SNR-aware fusion mechanism for robust AMC. In the proposed framework, CNN layers are employed to extract discriminative local spatial representations from raw *I/Q* samples, while BiLSTM layers model temporal dependencies inherent in communication signals. In addition, an auxiliary branch is introduced to explicitly encode SNR information, enabling the model to incorporate channel condition awareness into the learning process. The extracted signal features and SNR-aware representations are subsequently fused to improve classification robustness under varying noise conditions.

The core contributions of this work are summarized as follows:

- **Compact SNR-Aware Hybrid Architecture:** A compact CNN-BiLSTM framework with an independent SNR-processing branch is proposed for automatic modulation classification. The CNN layers extract local modulation-specific patterns from raw *I/Q* samples, while the BiLSTM layers capture forward and backward temporal dependencies in the signal sequence. Unlike conventional CNN-BiLSTM-based AMC models that rely only on waveform features, this architecture additionally incorporates an explicit SNR-aware pathway to provide channel-condition information to the classifier.
- **Explicit Feature-Level SNR Fusion:** The proposed architecture transforms the scalar SNR value into a low-dimensional embedding through a separate dense branch and concatenates it with the CNN–BiLSTM spatial-temporal feature representation before classification. This feature-level fusion enables the classifier to exploit both waveform structure and channel-quality information. As a result, the model can adjust its decision process according to the noise level, which is particularly useful in low-SNR regimes.
- **Benchmark-Based Experimental Validation:** The proposed method is evaluated on RadioML2016.10a and RadioML2016.10b, with detailed analysis across SNR regimes, modulation types, computational complexity, and comparison with recent state-of-the-art AMC methods. In addition to global accuracy metrics, the evaluation includes SNR-dependent performance curves, confusion matrices, and an ablation study.

These contributions are specifically intended to address limitations observed in recent state-of-the-art AMC models. CNN–LSTM and CNN–BiLSTM–based methods are effective for extracting spatial-temporal features from *I/Q* signals, but they generally learn the impact of channel quality implicitly from the waveform. Attention–based and Transformer–based models improve feature selection or global dependency modeling, but they may increase computational complexity and do not necessarily provide explicit channel-condition awareness. The proposed framework is therefore positioned as a compact SNR-aware alternative that combines spatial-temporal feature extraction with explicit feature-level fusion of channel-condition information.

The remainder of this paper is organized as follows. Section 2 reviews recent DL-based approaches for AMC and highlights the key challenges reported in the literature. Section 3 presents the system and signal model. Section 4 details the proposed hybrid architecture, the adopted datasets, and the preprocessing procedures. Section 5 reports the experimental results and provides an in-depth discussion of the model performance under varying SNR conditions. Finally, Section 6 concludes the paper and outlines directions for future research.

2. Related Work

Automatic Modulation Recognition (AMR) has transitioned from classical feature engineering to deep learning (DL) due to the demand for higher accuracy in complex channel conditions. The existing literature can be organized into three main architectural trends: spatial-temporal networks, attention/graph models, and global transformer or multi-domain fusion frameworks.

2.1 Spatial-Temporal Deep Learning Models

The integration of CNNs and recurrent neural networks (RNNs) established the baseline for modern DL-based AMR, focusing on extracting localized structural features alongside temporal sequence correlations.

Liu et al. [20] proposed an efficient DL model optimized for generalized AMR. The primary strength of this work lies in its highly optimized parameter architecture, which significantly lowers computational overhead, making it viable for constrained environments. However, its main limitation is a reduced capacity to handle non-linear channel impairments; the generalized nature of the model causes a sharp drop in classification accuracy when exposed to heavily faded or real-world non-Gaussian channels.

Addressing general radio frequency (RF) classification, Polak et al. [21] systematically benchmarked various classic DL architectures. The merit of this study is its comprehensive structural mapping, which provides clear guidelines on how depth and filter sizes affect raw *I/Q* signal processing. Nevertheless, a critical limitation of their approach is its strict reliance on standard grid-like topologies, which fail to scale efficiently and suffer from geometric computation bottlenecks as the temporal sequence length grows.

To retain complex-valued characteristics, Wu et al. [22] introduced a framework pairing complex-valued convolutions with an enhanced BiLSTM. Its major advantage is the preservation of phase relationships inherent in *I/Q* data, which considerably strengthens the feature representation of the network. Nonetheless, the sequential dependency of their enhanced BiLSTM block introduces substantial inference latency, creating a major bottleneck for real-time edge deployment, while remaining highly vulnerable to extreme noise at very low SNR levels.

2.2 Attention-Driven and Graph-Based Networks

To overcome the limitations of early spatial-temporal models—specifically their inability to selectively focus on crucial signal variations amidst dense noise—researchers introduced dynamic weighting and non-Euclidean relational modeling.

El-Haryqy et al. [23] integrated an additive attention mechanism with DL to focus on informative time-frequency segments. The strength of this design is its ability to dynamically weigh critical temporal bursts, which markedly improves classification thresholds in noisy environments. However, its vulnerability stems from the additive nature of the attention layer, which struggles to capture long-range global dependencies across the entire signal envelope, leaving the model sub-optimal against complex multi-path fading.

Shifting away from rigid grid structures, Zhang et al. [24] proposed an Adaptive Relational Graph Attention Network for RF proximity sensor signals. The core benefit of this approach is its capacity to extract non-local relational dependencies and subtle intra-class variations by treating signal elements as interconnected nodes. On the downside, its bottleneck is the immense computational complexity required to dynamically construct and update the graph topology at runtime, making it impractical for high-throughput, real-time wireless communication systems.

2.3 Transformer Frameworks and Multi-Source Feature Fusion

To bypass the local processing limitations of attention layers and the extreme computational cost of graph updates, recent paradigms have shifted toward Transformer architectures with global receptive fields and heterogeneous feature fusion.

Huo et al. [25] developed SMTrans, an efficient AMR network built on a scale-aware modulation transformer. Its main contribution is the use of a scale-aware mechanism that captures multi-scale global contexts without incurring the quadratic computational costs typical of standard self-attention. However, a notable drawback is that Transformers inherently lack inductive bias; consequently, SMTrans requires massive amounts of training data to converge and is highly prone to overfitting when applied to small-scale or highly corrupted datasets.

Concurrently, Cheng et al. [26] designed TriFusion-Lite, a lightweight network executing multi-domain fusion across temporal, frequency, and phase dimensions. The distinct advantage of this architecture is its lightweight footprint coupled with its ability to aggregate heterogeneous signal perspectives. Nevertheless, its main drawback is that the fusion is performed uniformly: the model does not differentiate between channel conditions and cannot exploit explicit external information such as an SNR estimate. As a result, its accuracy degrades abruptly when the operating environment experiences sudden noise drops.

2.4 Comparative Summary and Positioning of the Proposed Method

Recent state-of-the-art methods for AMC can be grouped into three main categories. CNN-LSTM/BiLSTM-based models combine convolutional local feature extraction with recurrent temporal modeling, which improves recognition performance compared with CNN-only or RNN-only architectures. Attention-based and Transformer-based models further improve feature selection and global dependency modeling, while multi-domain fusion approaches exploit complementary temporal, frequency, and phase representations. Most of these methods, however, rely mainly on raw *I/Q* samples and do not explicitly exploit external channel-condition information.

Table 1 summarizes the main characteristics, advantages, and limitations of representative recent AMC meth-

Study	Main architecture	Main advantages	Main limitations	Difference from proposed method
Liu et al. [27]	DCN–BiLSTM	Combines convolutional feature extraction and BiLSTM temporal modeling; effective for *I/Q*-based AMC.	Does not explicitly include an independent SNR-processing branch.	The proposed method adds explicit SNR-aware feature fusion.
Zhang et al. [16]	CNN–BiLSTM–DNN	Provides good recognition performance under low-SNR conditions using a CNN–BiLSTM–DNN structure.	Mainly relies on *I/Q* signal features; no separate SNR embedding branch is used.	The proposed method fuses CNN–BiLSTM features with SNR-based channel-condition information.
El-Haryqy et al. [23]	CNN–BiLSTM with additive attention	Attention mechanism improves the selection of informative temporal features.	The attention mechanism increases architectural complexity; no explicit SNR-input branch is included.	The proposed method uses a compact SNR-aware fusion branch instead of an attention mechanism.
SMTrans [25]	Scale-aware Transformer	Captures multi-scale global dependencies and improves feature representation.	Transformer–based models may require higher computational resources and larger datasets.	The proposed method keeps a compact CNN–BiLSTM backbone with explicit SNR conditioning.
TriFusion–Lite [26]	Temporal-frequency-phase fusion	Provides lightweight multi-domain feature fusion using temporal, frequency, and phase information.	Fusion is mainly signal-domain based and does not explicitly inject SNR metadata.	The proposed method introduces explicit channel-condition fusion through an SNR embedding.
Proposed CNN–BiLSTM–SNR	CNN–BiLSTM with separate SNR branch	Compact architecture; combines local, temporal, and SNR-aware channel information; improves robustness under varying SNR conditions.	Requires an available SNR value; robustness to SNR-estimation errors requires further validation.	N/A

Tab. 1. Comparative summary of recent AMC methods and the proposed framework.

ods. It shows that most existing models improve spatial-temporal feature extraction from *I/Q* samples, whereas few of them incorporate channel-condition information as a separate input stream. The proposed method differs from these approaches by introducing an independent SNR-processing branch and by performing explicit feature-level fusion between the CNN-BiLSTM waveform representation and the SNR-based channel-condition embedding.

3. System and Signal Model

To establish a rigorous framework for AMC, this section delineates the mathematical formulation of the transmitted RF signals, the physical phenomena governing the wireless propagation channel, and the multi-stream data representation ingested by the receiver.

3.1 Mathematical Modeling of the Received Signal

In a wireless communication system, the received signal captured by the receiving antenna can be described by its continuous-time complex baseband equivalent, expressed as:

$$r(t) = s(t; m) * h(t) + n(t) \tag{1}$$

where,

- $S(t; m)$ represents the baseband signal generated by the transmitter using the m-th modulation scheme out of M possible configurations, with $m \in \{1, 2, \ldots, M\}$.
- $h(t)$ denotes the time-varying complex impulse response of the wireless channel, which introduces physical propagation impairments such as multipath fading, Doppler shifts, and phase offsets.
- $n(t)$ represents the AWGN introduced by the receiver front-end electronics, modeled as a zero-mean complex Gaussian random process with variance σ_n^2.

3.2 Signal Digitization and Observation Window

Before any DL or statistical feature extraction is performed, the continuous-time received waveform $r(t)$ is first processed by an Analog-to-Digital Converter (ADC) operating at a uniform sampling frequency $f_s = \frac{1}{T_s}$, where T_s denotes the sampling interval. The digitized signal is subsequently segmented into discrete observation blocks. Each signal instance is captured within a finite temporal window consisting of a fixed number of N consecutive samples ($N = 128$ in standard benchmark datasets such as RadioML 2016.10a).

The resulting discrete-time signal is expressed using the standard orthogonal Cartesian *I/Q* representation as:

$$r[n] = x_I[n] + j\, x_Q[n], \quad n = 0, 1, \ldots, N-1 \tag{2}$$

where $x_I[n]$ and $x_Q[n]$ denote the I and Q components, respectively. Together, these components preserve both the amplitude envelope and the phase variations introduced by the original modulation process.

3.3 Multi-Stream Input Representation

In cognitive radio and AMC scenarios, relying solely on raw waveform samples may limit classification performance, particularly under severe channel impairments and low-SNR conditions. To enhance the discriminative capability of the proposed network, we define two complementary input streams in the proposed framework: the behavioral waveform stream ($\mathbf{X}_{\text{sig}}$) is represented as a two-dimensional *I/Q* matrix containing the raw temporal dynamics of the received signal:

$$\mathbf{X}_{\text{sig}} = \begin{bmatrix} x_{\text{I}}[0] & x_{\text{I}}[1] & \cdots & x_{\text{I}}[N-1] \\ x_{\text{Q}}[0] & x_{\text{Q}}[1] & \cdots & x_{\text{Q}}[N-1] \end{bmatrix} \in \mathbb{R}^{2\times N}. \quad (3)$$

The environmental contextual stream is represented by a scalar SNR value associated with each signal sample. In the current RadioML-based experimental setup, this value corresponds to the dataset-provided SNR label and is used as an ideal channel-condition indicator:

$$x_{\text{SNR}} = \gamma_{\text{dB}} \in \mathbb{R}^{1\times 1}. \quad (4)$$

By formally defining these two independent input domains, the proposed framework motivates the use of a dual-branch DL architecture, in which the waveform stream and the SNR stream are processed separately before feature-level fusion, as presented in the subsequent section.

4. Dataset and Model Description

4.1 Model Architecture

The proposed model is a hybrid CNN-BiLSTM architecture with explicit SNR-aware feature fusion, designed for robust AMC under varying channel conditions. The input consists of raw complex baseband *I/Q* signals represented by 128 time-domain samples with I and Q components, together with an associated scalar SNR value for each sample.

Initially, the *I/Q* samples are processed by a one-dimensional convolutional feature extraction block, which treats the I and Q components as separate input channels, enabling the network to learn local modulation-specific patterns directly from the raw waveform. This block comprises successive Conv1D layers with an increasing number of filters, each followed by batch normalization and ReLU activation. Max-pooling operations are interleaved between layers to progressively reduce the temporal resolution while enhancing the discriminative feature representation. The scalar SNR input corresponds to the dataset-provided SNR label, which is treated as an ideal channel-condition indicator. This assumption is further clarified in the dataset description section.

The resulting high-level feature maps are then reshaped and passed to a two-layer BiLSTM. This component captures both forward and backward temporal dependencies within the signal sequence, allowing the model to learn long-range temporal correlations that are important for distinguishing modulation schemes with similar instantaneous characteristics. The output of the BiLSTM at the last time step is used as a compact temporal representation of the signal and is projected through a fully connected layer to produce a fixed-length embedding.

In parallel, the scalar SNR value is processed by a dedicated lightweight fully connected branch, which transforms it into a low-dimensional embedding. This explicit pathway enables the model to incorporate channel-quality awareness into the decision-making process. The two embeddings—one representing learned signal features and the other reflecting channel noise—are then concatenated and fed into a final classification module composed of fully connected layers with nonlinear activation functions and dropout regularization.

The feature-level fusion strategy allows the classifier to dynamically adapt its decision boundaries based on both signal structure and noise level, thereby improving robustness, particularly in low-SNR regimes. By jointly leveraging convolutional spatial feature extraction, bidirectional temporal modeling, and explicit SNR conditioning, the proposed CNN-BiLSTM architecture with SNR fusion provides a compact and flexible framework for AMC under varying channel conditions.

The novelty of the proposed architecture does not lie in the individual CNN or BiLSTM components, which have already been used in AMC, but in the explicit integration of channel-condition information within a compact dual-branch framework. Unlike conventional CNN-BiLSTM classifiers that process only raw *I/Q* samples and learn noise-dependent behavior implicitly, the proposed model includes an independent SNR branch that maps the scalar SNR value to a trainable low-dimensional embedding. This embedding is concatenated with the CNN-BiLSTM spatiotemporal representation at the feature level. Consequently, the classification module receives both modulation-dependent waveform information and an explicit indication of waveform reliability. This formulation enables the network to learn SNR-conditioned decision boundaries while avoiding the computational overhead associated with attention, graph, or Transformer modules. The architecture therefore combines hierarchical local feature extraction, bidirectional temporal modeling, explicit channel awareness, and compact computational complexity within a unified AMC framework. In the current benchmark evaluation, the dataset-provided SNR is used as an ideal channel-condition indicator. Thus, the results represent an upper-bound assessment of SNR-aware fusion, and robustness to practical SNR-estimation errors remains to be evaluated.

Figure 1 illustrates the overall architecture of the proposed model.

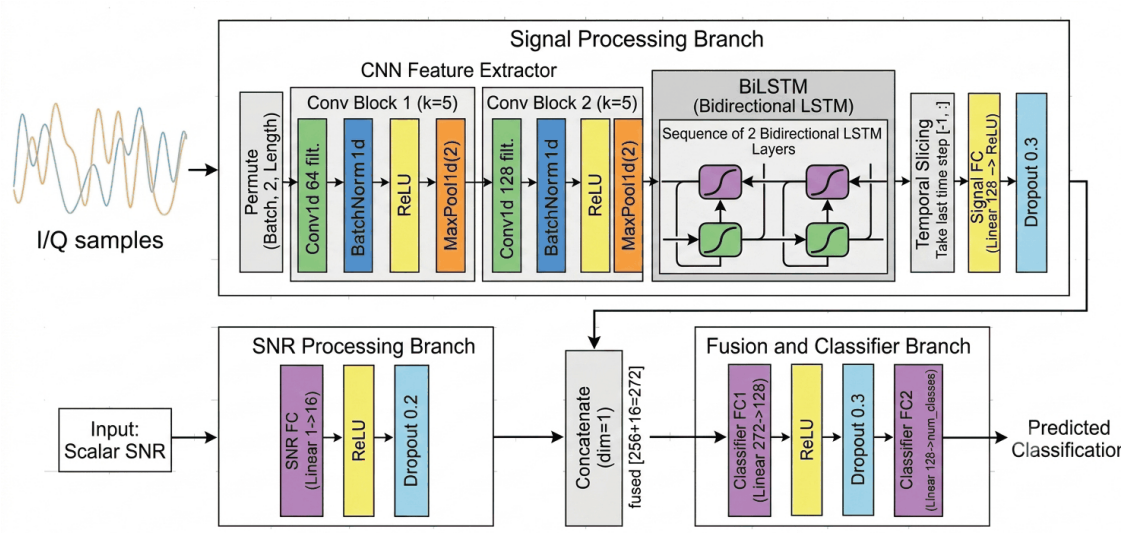


Fig. 1. Overall architecture of the proposed model.

4.2 Datasets Description

The experiments are conducted using the RadioML2016.10a and RadioML2016.10b datasets [8], which are widely recognized benchmarks for AMC and AMR in wireless communication systems. Although RadioML datasets include simulated channel impairments such as AWGN, carrier frequency offset, sample-rate offset, and multipath fading, they remain synthetic benchmark datasets. Consequently, they are useful for reproducible AMC evaluation under controlled conditions, but they do not capture all real over-the-air propagation effects, hardware impairments, or receiver-side uncertainties.

Both datasets are synthetically generated and consist of complex baseband *I/Q* signal samples, enabling end-to-end machine learning and DL approaches to operate directly on raw signals without handcrafted features. Each signal instance is represented by 128 time-domain samples with two channels corresponding to the I and Q components, facilitating efficient temporal feature extraction and sequence modeling.

The datasets encompass a diverse set of digital and analog modulation schemes, forming a challenging multi-class classification problem. In particular, both datasets share the same set of eight digital modulation formats: BPSK, QPSK, 8PSK, 16QAM, 64QAM, PAM4, GFSK, and CPFSK, while they differ in their analog modulation content: RadioML2016.10a includes three analog modulations (AM-DSB, AM-SSB, and WBFM), whereas RadioML2016.10b includes only two analog modulations (AM-DSB and WBFM). Each signal sample has a data length of 128, represented as a 2×128 *I/Q* data matrix, with a sampling frequency of 200 kHz.

To approximate practical wireless propagation effects in a controlled simulation setting, both datasets incorporate several channel impairments [27], including AWGN with a noise standard deviation of $\sqrt{10\,\mathrm{e}^{-\mathrm{SNR}/10}}$, carrier frequency offset with a standard deviation of 0.01 Hz and a maximum offset of 500 Hz, sample rate offset with a standard deviation of 0.01 Hz and a maximum offset of 50 Hz, and multipath fading effects modeled using eight sinusoids with a maximum Doppler frequency of 1 Hz, following a Rician fading model with a K-factor of 4, a power delay profile using fractional sample delays of $[0.0, 0.9, 1.7]$ with corresponding magnitudes of $[1, 0.8, 0.3]$, and an interpolation filter length of 8.

The signal samples span a wide range of SNR levels from −20 dB to +18 dB in 2 dB increments, allowing comprehensive evaluation across highly noisy to near-ideal channel conditions. The SNR values used in this work correspond to the SNR labels provided by the RadioML datasets. Since RadioML2016.10a and RadioML2016.10b are synthetically generated using GNU Radio channel simulation blocks, the SNR is a controlled simulation parameter associated with the AWGN level during dataset generation. Therefore, the SNR input is used as an ideal channel-condition indicator to evaluate the potential benefit of SNR-aware feature fusion. This setting should be interpreted as an upper-bound evaluation rather than a fully realistic receiver-side SNR estimation scenario. In practical AMC receivers, the SNR information would be provided by an SNR-estimation or channel-quality estimation module and may therefore be affected by estimation errors. With approximately 1.2 million well-balanced labeled samples in RadioML2016.10b and comparable diversity in RadioML2016.10a, these datasets are well suited for reproducible benchmark evaluation of low-SNR performance under controlled synthetic channel conditions.

Figure 2 illustrates representative time-domain *I/Q* samples for the modulation classes in the RadioML2016.10a dataset at an SNR of 6 dB. Different modulation types produce clearly different temporal patterns in their I and Q components under the same SNR condition, which illustrates the nature of the input data and motivates the use of a deep learning model capable of extracting discriminative features directly from raw *I/Q* sequences.

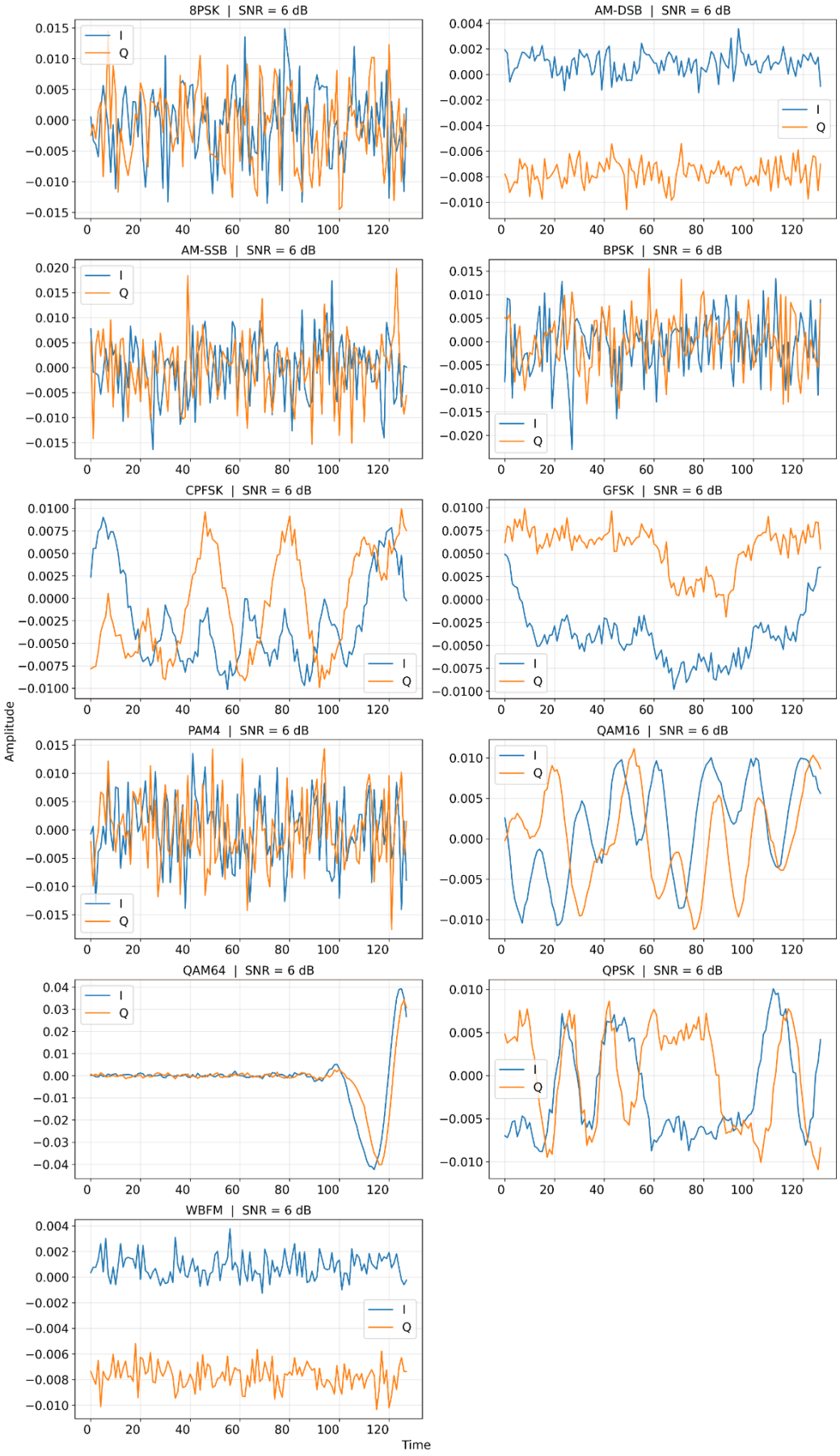


Fig. 2. Time-domain I and Q components of different modulation signals at an SNR of 6 dB.

Although RadioML2016.10a and RadioML2016.10b provide reproducible benchmarks and include simulated effects such as additive noise, carrier-frequency offset, sample-rate offset, multipath propagation, and fading, both datasets remain synthetically generated. Consequently, they do not fully capture practical transmitter, propagation, and receiver effects, including power-amplifier nonlinearity, oscillator phase noise, *I/Q* imbalance, analog filtering, antenna characteristics, receiver calibration errors, quantization effects, and uncontrolled interference. The present evaluation should therefore be interpreted as a controlled benchmark assessment rather than complete validation under real operating conditions. Moreover, although the datasets contain several representative digital modulation families, the number and diversity of high-order digital modulation formats remain limited. Further evaluation will consider datasets containing a broader range of digital modulation schemes and measured over-the-air signals. In such a setting, the proposed SNR branch will receive an estimated rather than ideal SNR value, allowing the robustness of the feature-fusion mechanism to be evaluated under realistic SNR-estimation errors.

4.3 Data Preprocessing

The raw *I/Q* data are first preprocessed to ensure compatibility with the proposed DL architecture. Each signal sample is reshaped into a standardized format of 128×2, where 128 denotes the temporal dimension and the two channels correspond to the I and Q components, which provides a consistent input structure across all samples. The modulation labels are then converted into numerical representations using one-hot encoding to support multi-class classification, and the dataset is divided into training, validation, and test subsets using a stratified sampling strategy that ensures proportional representation of each modulation. Finally, the SNR values are standardized to zero mean and unit variance, which improves numerical stability during training of the SNR-aware branch.

5. Experiments and Results Analysis

5.1 Experimental Setup

The experimental setup is implemented using the PyTorch DL framework, which enables efficient design and training of the proposed hybrid CNN-BiLSTM-SNR architecture. All experiments are conducted on a workstation equipped with an NVIDIA GeForce RTX 4070 SUPER GPU, while CPU execution is used for compatibility and debugging purposes. The software environment is based on Anaconda with Python 3.8.20, PyTorch 2.4.1+cu118 (CUDA 11.8), and cuDNN 9.1.0. The evaluation relies on standard classification metrics, namely accuracy, precision, recall, and F1-score. The hyperparameters chosen are summarized in Tab. 2.

Category	Hyperparameter Value
Data	Datasets / Classes / Input RadioML 2016.10a / 11 / (128, 2) RadioML 2016.10b / 10 / (128, 2)
Split	Train / Val / Test / Seed 70% / 15% / 15% / 42
Training	Batch Size / Epochs 256 / 50
Optimizer	Adam / LR / Weight Decay Adam / 1×10^{-3} / 1×10^{-5}
Scheduler	ReduceLROnPlateau factor = 0.5, patience = 3
CNN Conv Layers	3 layers: $2 \rightarrow 64 \rightarrow 128 \rightarrow 128$ Kernels: 5, 5, 3
Activation / Norm / Pool	ReLU / BatchNorm1d / MaxPool1d(2)
BiLSTM	Hidden Size / Layers / Bidirectional 128 / 2 / True
FC Layers	Signal: $256 \rightarrow 128$ (Dropout 0.3) SNR: $1 \rightarrow 16$ (Dropout 0.2) Classifier: $144 \rightarrow 128 \rightarrow 11$ (Dropout 0.3)
Output	Label Encoding One-Hot (batch, 11)

Tab. 2. Summary of the hyperparameters of the proposed CNN-BiLSTM model with a separate SNR branch.

To support the research community and ensure reproducibility, the source code for the proposed CNN-BiLSTM-SNR framework is publicly available and accessible at https://github.com/HichamUsto/CNN-BiLSTM-SNR.

5.2 Experiment Results

To evaluate the effectiveness of the proposed architecture under the considered RadioML benchmark conditions, a comprehensive experimental evaluation is conducted. To this end, this section presents and discusses the results obtained on RadioML2016.10a and RadioML2016.10b datasets, including overall classification accuracy curves, SNR-specific confusion matrices, and comparative analyses against state-of-the-art methods.

5.2.1 RadioML2016.10a

Figures 3 and 4 present the evolution of the training and validation accuracy and loss, respectively, as a function of the number of epochs on the RadioML2016.10a dataset. As training progresses, the accuracy increases while the loss decreases, indicating that the model learns discriminative features and minimizes the objective function. The validation curves closely follow the training curves, with only limited fluctuations, and no divergence between them is observed. This behavior indicates stable convergence and no significant overfitting under the adopted benchmark protocol.

The confusion matrix depicted by Fig. 5 shows that the classifier achieves strong overall performance, with the majority of samples correctly classified along the main diagonal. The highest correct classification rates are obtained for BPSK, QAM16, QAM64, AM-DSB, and PAM4.

The remaining errors mainly occur between modulation types with similar signal characteristics, such as AM-DSB versus AM-SSB, QAM16 versus QAM64, and CPFSK versus GFSK. The main exception is WBFM: 1,653 of its 3,000 frames are assigned to AM-DSB, leaving only 819 correct decisions. Overall, the matrix confirms the effectiveness of the proposed model in discriminating between the digital modulation schemes, the residual ambiguity being confined to the analog pair AM-DSB/WBFM.

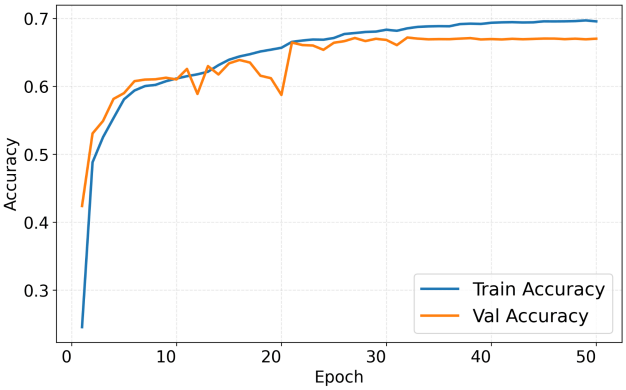


Fig. 3. Training and validation accuracy versus epoch (RadioML2016.10a).

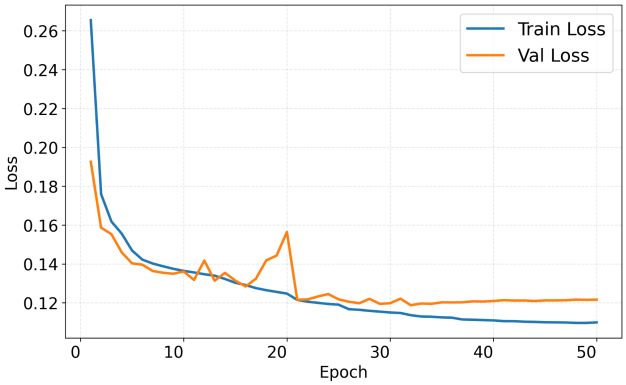


Fig. 4. Training and validation loss versus epoch (RadioML2016.10a).

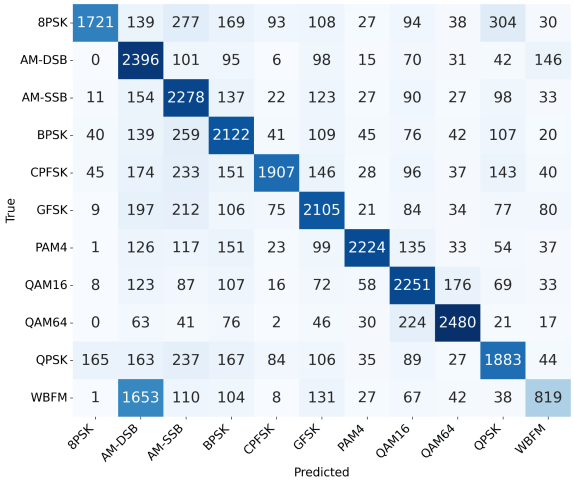


Fig. 5. Global confusion matrix (RadioML2016.10a).

Figure 6 illustrates the model's performance metrics (acc., prec., recall, and F1-score) across a range of SNR. As expected, performance degrades below −12 dB, where all four scores fall to approximately 0.10, that is, the level of a random decision over eleven classes. The scores then improve almost linearly between −12 dB and 0 dB, before stabilising on a plateau close to 0.93 for all positive SNR values. The close alignment of all four metrics indicates that the classifier is not biased towards a particular class and that precision is not traded against recall.

Figure 7 reports the confusion matrices obtained at selected SNR levels ranging from −6 dB to 6 dB. At −6 dB (Acc = 0.67), errors are concentrated between the analog modulations (AM-DSB and WBFM) and between the phase-shift keying schemes (8PSK and QPSK). As the SNR increases to 6 dB (Acc = 0.93), the diagonal becomes dominant and most digital schemes are classified almost perfectly. A persistent confusion nevertheless remains between AM-DSB and WBFM at high SNR, and it is one-directional: at 6 dB, 91 of the 149 WBFM frames are predicted as AM-DSB, whereas AM-DSB is almost never assigned to WBFM. Because it does not vanish as the noise decreases, this error reflects the limited discriminative information carried by the WBFM envelope over a 128-sample window rather than a noise-related limitation.

5.2.2 RadioML2016.10b

Figures 8 and 9 present the training and validation performance of the model on the RadioML2016.10b dataset in terms of accuracy and loss, respectively. The training accuracy rises steadily and the validation accuracy follows the same trajectory, although pronounced fluctuations appear over the first ten epochs, during which the validation loss temporarily rises to about 0.22. Both curves then flatten after approximately thirty epochs, the training and validation accuracies converging towards 0.69 and the corresponding losses towards 0.12. The absence of a significant gap between training and validation curves in both accuracy and loss indicates stable learning behaviour and limited overfitting on this dataset.

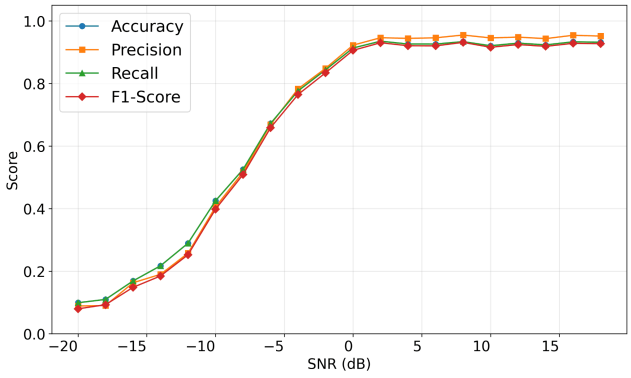


Fig. 6. Performance metrics versus SNR (combined view), RadioML2016.10a.

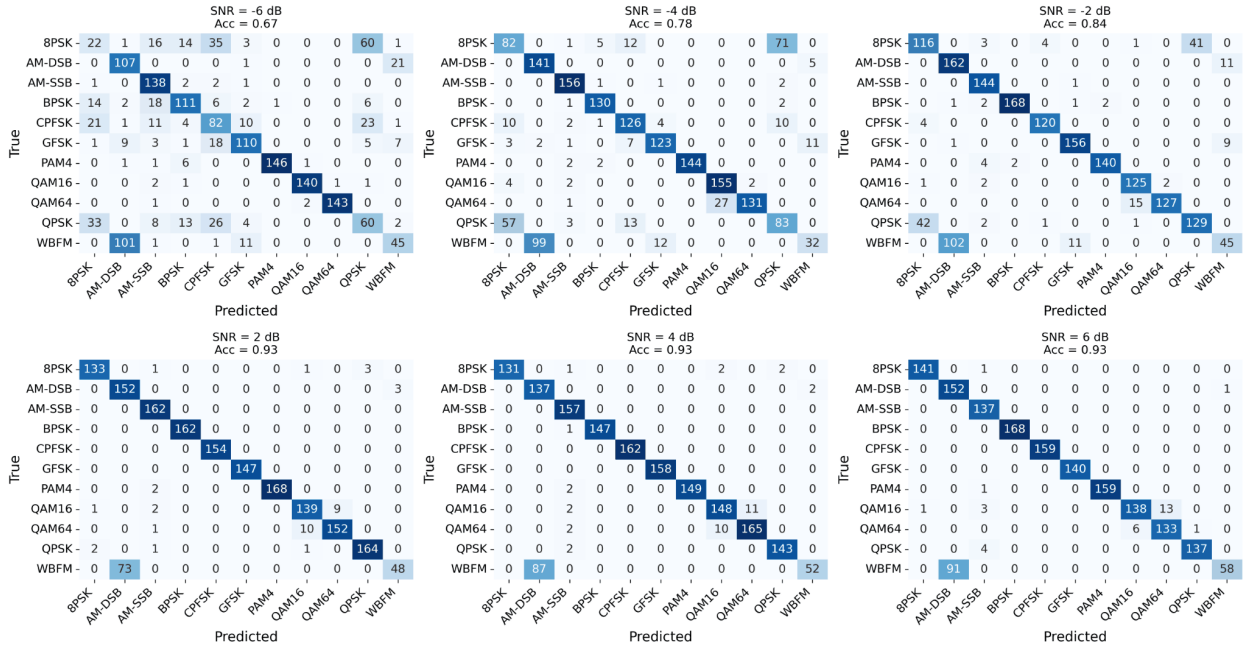


Fig. 7. Confusion matrices at different SNR levels (RadioML2016.10a).

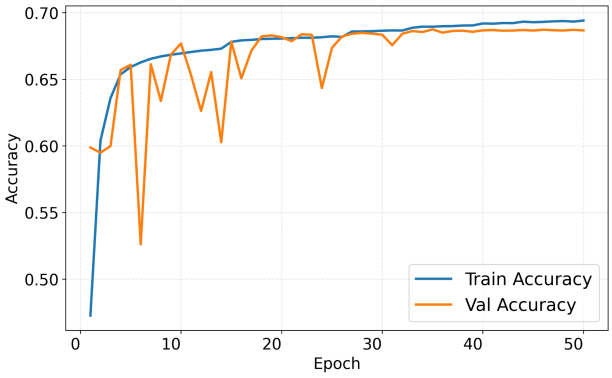


Fig. 8. Training and validation accuracy versus epoch (RadioML2016.10b).

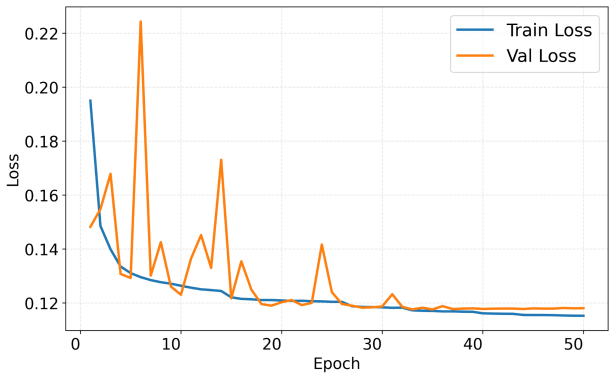


Fig. 9. Training and validation loss epoch (RadioML2016.10b).

The validation accuracy and validation loss exhibit pronounced fluctuations during approximately the first ten epochs. These variations are mainly associated with the transient optimization stage, during which the network parameters and classification boundaries change rapidly under mini-batch training. The validation set contains ten modulation classes distributed across SNR levels from –20 dB to +18 dB, resulting in a heterogeneous mixture of relatively easy high-SNR observations and severely degraded low-SNR observations. During early training, small changes in the learned representation can therefore alter the predictions of many samples located close to the decision boundaries. The validation loss shows stronger fluctuations because cross-entropy depends not only on the predicted class but also on prediction confidence. Consequently, a small number of highly confident incorrect predictions can produce a visible loss increase without an equivalent decrease in accuracy. Differences between the mini-batch statistics used by batch normalization during training and the accumulated statistics used during validation may also contribute to the early variations. Nevertheless, these fluctuations decrease as training progresses. After approximately 30 epochs, the training and validation accuracies converge toward 0.69, and the corresponding losses stabilize near 0.12. The absence of a persistent divergence between the training and validation curves suggests that the observed fluctuations represent an early optimization transient rather than progressive overfitting.

Figure 10 presents the cumulative confusion matrix of the model across all tested SNR environments. The diagonal elements show that the network successfully learns discriminative features for the majority of digital modulations, especially QAM64, QAM16, and PAM4. However, two prominent error patterns emerge: first, a one-directional confusion of WBFM with AM-DSB, since 9,950 of the 18,000 WBFM frames are predicted as AM-DSB whereas only 684 AM-DSB frames are assigned to WBFM; second, a mutual overlap between 8PSK and QPSK, where 2,396 QPSK frames are predicted as 8PSK and 1,224 8PSK frames as QPSK. These

confusions reflect the geometric similarity of the two phase-shift keying constellations and the limited discriminative information carried by the WBFM envelope over a 128-sample window.

Figure 11 presents a series of confusion matrices illustrating the modulation classification performance across different SNR levels (from −6 dB to 6 dB). At low SNR (e.g. −6 dB), the matrices exhibit significant off-diagonal elements indicating frequent misclassifications, particularly between the phase- and frequency-modulated schemes 8PSK, QPSK, and CPFSK, and between WBFM and AM-DSB, where 610 of the 936 WBFM frames are assigned to AM-DSB. As the SNR increases, the diagonal dominance becomes progressively stronger, reflecting improved class separability.

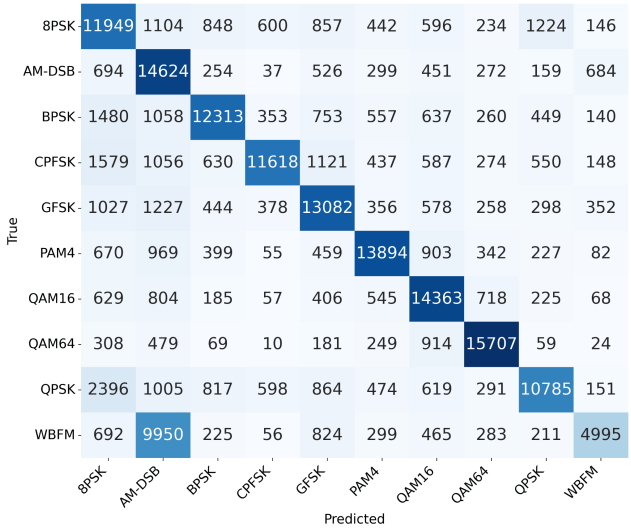


Fig. 10. Global confusion matrix (RadioML2016.10b).

From −2 dB onward, most modulation types are correctly identified with high accuracy, and at higher SNRs (2–6 dB) the confusion matrices are nearly diagonal for the eight digital modulations, the overall accuracy saturating at 0.93. The residual error is almost entirely due to WBFM, of which 501 frames are still predicted as AM-DSB at 6 dB against 374 correct decisions, which explains why the accuracy no longer improves once the noise becomes negligible.

Figure 12 illustrates the evolution of classification performance metrics as a function of SNR. It jointly reports accuracy, precision, recall, and F1-score, all of which follow the same trend. Below −12 dB the four metrics remain close to 0.12, that is, near the level of a random decision over ten classes, whereas a rapid improvement is observed as the SNR moves toward moderate values. Beyond approximately 0 dB, all metrics saturate slightly above 0.93 rather than approaching unity, the residual gap being caused by the WBFM frames absorbed by AM-DSB.

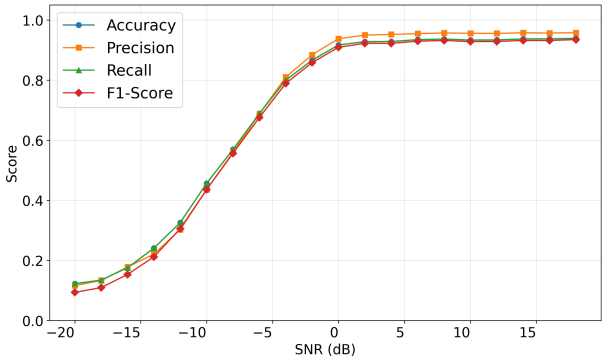


Fig. 12. Performance metrics versus SNR (combined view), RadioML2016.10b.

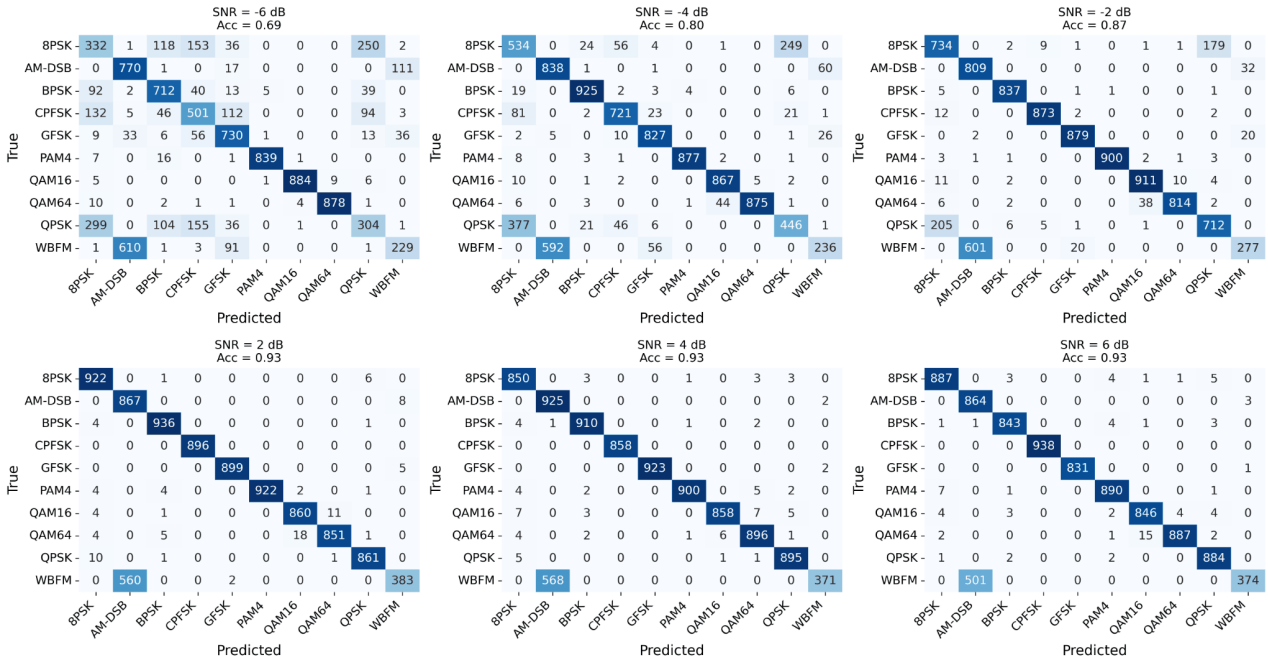


Fig. 11. Confusion matrices at different SNR levels (RadioML2016.10b).

Dataset	Model	Acc.	F1–score	Prec.
RadioML 2016.10a	CNN-only	0.5783	0.5546	0.5991
	BiLSTM-only	0.0910	0.0152	0.0083
	CNN-BiLSTM	0.6195	0.5974	0.6290
	CNN-BiLSTM + SNR	**0.6723**	**0.6571**	**0.6732**
RadioML 2016.10b	CNN-only	0.6361	0.6232	0.6587
	BiLSTM-only	0.1000	0.0182	0.0100
	CNN-BiLSTM	0.6482	0.6339	0.6665
	CNN-BiLSTM + SNR	**0.6852**	**0.6729**	**0.6930**

Tab. 3. Ablation study results of the proposed model on the RadioML2016.10a and RadioML2016.10b datasets.

5.3 Ablation Study: Evaluating Component Contributions

To further validate the effectiveness of the proposed approach, an ablation study was conducted to evaluate the contribution of each architectural component by comparing four model variants, namely a CNN-only model, a BiLSTM-only model, a CNN-BiLSTM model without SNR, and the complete CNN-BiLSTM architecture augmented with the SNR branch, with all experiments performed under identical conditions to ensure a fair and consistent comparison.

The results reported in Tab. 3 clearly demonstrate that the full CNN-BiLSTM model with the SNR branch achieves the best overall performance on both RadioML2016.10a and RadioML2016.10b datasets, reaching accuracies of 67.23% and 68.52%, respectively, which highlights the benefit of jointly leveraging local feature extraction through CNNs, temporal dependency modeling through BiLSTMs, and SNR-aware information to enhance classification robustness.

In contrast, the CNN-only model, although effective in capturing local discriminative patterns, yields lower accuracies of 57.83% on RadioML2016.10a and 63.61% on RadioML2016.10b, indicating its limited capability in modeling long-term temporal dependencies inherent in modulation signals.

The BiLSTM-only configuration exhibits the weakest performance, with accuracies of 9.10% and 10.00% on the two datasets, demonstrating that temporal modeling alone is insufficient without effective low-level feature extraction.

The CNN-BiLSTM model without the SNR branch achieves accuracies of 61.95% and 64.82% on RadioML2016.10a and RadioML2016.10b, respectively. This intermediate performance shows that combining CNN and BiLSTM already improves classification compared with using each component separately. However, the additional improvement obtained by the CNN-BiLSTM-SNR model confirms the specific contribution of the SNR-aware branch.

Consistent trends are observed across the F1 score and precision metrics, where the CNN-BiLSTM model with the SNR branch consistently achieves the highest values, including precision scores of 0.6732 and 0.6930 for RadioML2016.10a and RadioML2016.10b, respectively.

Overall, this ablation study indicates that the integration of CNNs, BiLSTMs, and SNR-aware information improves modulation classification performance, particularly across varying noise conditions.

Although the F1-scores of 0.6571 and 0.6729 were obtained by the complete model on RadioML2016.10a and RadioML2016.10b, respectively, may appear moderate; they are calculated over the entire SNR interval from –20 dB to +18 dB. At the lowest SNR levels, modulation-dependent structures are heavily masked by noise, leading to reduced class-wise precision and recall. The multiclass setting also includes modulation families with similar signal characteristics, such as 16QAM and 64QAM or QPSK and 8PSK, which remain difficult to distinguish under severe noise. Therefore, low-SNR classification errors have a substantial influence on the overall F1-score. Nevertheless, the ablation results demonstrate a consistent benefit from the proposed architecture. Relative to the waveform-only CNN-BiLSTM baseline, incorporating the SNR branch increases the F1-score from 0.5974 to 0.6571 on RadioML2016.10a and from 0.6339 to 0.6729 on RadioML2016.10b, corresponding to absolute improvements of 5.97 and 3.90 percentage points, respectively. Moreover, the F1-scores remain close to their corresponding overall accuracies, indicating that the performance improvement is jointly reflected in precision and recall rather than being restricted to a small subset of classes. However, these absolute values confirm that classification at extremely low SNR remains challenging and motivate future work on more robust feature extraction, channel-aware learning, and realistic channel evaluation.

5.4 Computational complexity comparison

To ensure a fair interpretation of Tab. 4, the inference-time values reported for the baseline models are taken from previously published studies [23] and may have been measured under different hardware, software, batch-size, and timing conditions. Therefore, Table 4 should be interpreted as a literature-based computational-complexity comparison rather than a fully controlled head-to-head inference-time benchmark. Definitive conclusions regarding inference-time superiority would require all models to be reimplemented and evaluated under the same execution protocol.

As summarized in Tab. 4, the proposed CNN-BiLSTM-Separate SNR model exhibits a favorable computational efficiency compared with several state-of-the-art baseline architectures, including ResNet, Transformer, CVCNN, PET-CGDNN, MCLDNN, DAE, MAMCA, and ICRNN. The comparison is conducted in terms of FLOPs, number of parameters, and inference time, which are useful indicators for assessing the potential suitability of the model for resource-constrained and edge-oriented AMC applications.

With only 4.63 MFLOPs, the proposed model is significantly less computationally demanding than heavy architectures such as ResNet (11,200 MFLOPs) and CVCNN (3,100 MFLOPs), while also outperforming PET-CGDNN

Model	FLOPs (M)	Parameters (M)	Inference Time (s)
ResNet [28]	11,200	53.8	0.909
Transformer [29]	408	0.0384	7.89
CVCNN [29]	3,100	68.1	9.22
PET-CGDNN [32]	86.9	79.2	2.29
MCLDNN [31]	1,170	406	4.45
DAE [30]	0.536	0.149	11.4
MAMCA [29]	73.9	16.8	0.246
ICRNNA [23]	48.42	0.79	0.2313
CNN-BiLSTM-Separate SNR (Ours)	4.63	0.804	0.0029

Tab. 4. Comparison of computational efficiency.

Dataset	Model	Avg. Acc.			Overall Accuracy (%)	Min Acc.	Max Acc.
		−20 ~ −8 dB	−6 ~ 4 dB	6 ~ 18 dB			
A	SMTrans [25]	17.95	81.47	**92.99**	63.27	9.52	93.34
B		21.30	82.97	**93.77**	65.17	10.62	93.88
A	MCLDNN [31]	17.09	79.29	92.30	62.08	9.25	92.71
B		20.65	81.97	93.57	64.57	10.81	93.70
A	PET-CGDNN [32]	17.35	76.20	90.78	60.71	9.40	91.07
B		20.57	80.59	93.19	63.99	10.49	93.34
A	AMC-Net [33]	17.33	78.70	90.24	61.26	9.32	90.64
B		20.61	83.48	93.53	65.00	10.60	93.65
A	FEA-T [34]	17.14	76.09	88.68	59.86	9.42	89.08
B		20.67	81.75	93.43	64.46	10.49	93.56
A	ICRNNA [23]	–	–	–	63.24	–	–
B		–	–	–	65.39	–	–
A	CBADNN [22]	–	–	–	64.02	**9.94**	**93.66**
B		–	–	–	65.50	10.89	94.20
A	**CNN-BiLSTM-SNR (Ours)**	**26.20**	**84.42**	92.82	**67.23**	9.92	93.49
B		**28.93**	**85.46**	93.58	**68.52**	**12.26**	**93.91**

Tab. 5. Comparison of classification performance with AMR models on different datasets (A: RadioML2016.10a, B: RadioML2016.10b).

and MCLDNN. Moreover, its compact design requires just 0.804 million parameters, which is orders of magnitude smaller than MCLDNN (406 million) and ResNet (53.8 million), and comparable to lightweight models such as ICRNN.

In terms of inference speed, CNN-BiLSTM-Separate SNR achieves a low inference time of 0.0029 s, which is faster than all competing models, including Transformer (7.89 s) and CVCNN (9.22 s). Although ultra-lightweight approaches such as DAE exhibit lower FLOPs and parameter counts, they suffer from considerably higher inference latency and reduced robustness.

Overall, the proposed model achieves a favorable trade-off between computational efficiency and classification performance under the evaluated benchmark conditions, suggesting its promising potential for resource-constrained AMC applications. However, deployment-oriented claims require further validation on real RF signals, independent channel conditions, hardware impairments, and practical receiver-side uncertainties.

5.5 Comparative study

Table 5 provides a comprehensive comparative evaluation of AMR models on the widely used RadioML2016.10a (A) and RadioML2016.10b (B) datasets, which differ mainly in channel conditions and impairment severity. The comparison includes several representative DL-based AMR architectures, namely SMTrans, MCLDNN, PET-CGDNN, AMC-Net, FEA-T, ICRNNA, CBADNN, and the proposed CNN-BiLSTM with separate SNR processing. Performance is reported in terms of average accuracy across SNR ranges, overall accuracy, as well as minimum and maximum accuracies, offering a detailed view of robustness under varying noise conditions.

The results are further stratified across three SNR regimes: low SNR (−20 to −8 dB), medium SNR (−6 to 4 dB), and high SNR (6 to 18 dB). In the low-SNR region, all models experience significant performance degradation due to severe noise and channel distortion; however, the proposed method exhibits noticeably higher average accuracy than competing approaches, indicating improved resilience under challenging low-SNR benchmark conditions. In the medium-SNR range, which is relevant to many communication scenarios, the proposed CNN-BiLSTM-SNR model consistently outperforms existing methods on both datasets, achieving the highest overall accuracy and demonstrating better discrimination among closely related modulation schemes. At high SNR values (6 to 18 dB), most models converge toward strong performance. While the proposed approach maintains a highly competitive accuracy (92.82% and 93.58% on datasets A and B, respectively), it is slightly outperformed by

the SMTrans model [25], which achieves the highest average accuracies in this regime (92.99% and 93.77%).

A key observation from the table is the consistent performance gain on RadioML2016.10b, which is generally more challenging than RadioML2016.10a. The proposed model achieves superior overall and minimum accuracies on dataset B, indicating improved performance under the channel impairments represented in the RadioML2016.10b benchmark. This improvement is attributed to the explicit separation and exploitation of SNR information, which allows the network to adapt its feature learning strategy according to channel quality rather than relying solely on raw waveform representations.

Overall, the comparative study indicates that the CNN-BiLSTM model with separate SNR input achieves competitive and, in several cases, improved performance compared with state-of-the-art AMR models across the considered evaluation criteria and SNR ranges. These results support the effectiveness of incorporating channel-condition information alongside temporal waveform features under RadioML benchmark conditions. Nevertheless, further validation is required before drawing definitive conclusions regarding performance in real over-the-air wireless environments.

6. Conclusion

In this paper, we proposed a compact hybrid DL architecture for AMC that combines hierarchical 1D CNN local feature extraction from raw *I/Q* sequences with BiLSTM temporal modeling and an independent SNR branch fused at the feature level with the waveform representation. This explicit channel-condition fusion, rather than the CNN–BiLSTM backbone alone, constitutes the methodological contribution: the ablation results attribute to it gains of 5.28 and 3.70 percentage points in overall accuracy on RadioML2016.10a and RadioML2016.10b, with up to an 8.25 percentage-point gain in the low-SNR regime, at a cost of only 4.63 MFLOPs. These results are bounded by the experimental setting: the SNR input is the dataset-provided label rather than a receiver-side estimate, and the inference-time comparison is literature-based, so the computational advantage remains indicative.

Future work will therefore explore attention mechanisms to better resolve ambiguities between similar modulation schemes such as AM-DSB and WBFM, along with quantization for embedded deployment. The evaluation will also be extended to a wider range of digital modulation formats, including additional high-order PSK, QAM, APSK, and multicarrier signals. Particular attention will be given to over-the-air measurements containing transmitter and receiver hardware impairments. Another important direction will be to replace the ideal dataset-provided SNR labels with practical SNR estimates and evaluate the sensitivity of the proposed fusion mechanism to estimation bias and uncertainty.

References

[1] HAYKIN, S. Cognitive radio: brain-empowered wireless communications. *IEEE Journal on Selected Areas in Communications*, 2005, vol. 23, no. 2, p. 201–220. DOI: 10.1109/JSAC.2004.839380

[2] GARG, V. *Wireless Communications & Networking*. Amsterdam: Elsevier, 2010. ISBN: 9780123735805

[3] SAAD, W., BENNIS, M., CHEN, M. A vision of 6G wireless systems: applications, trends, technologies, and open research problems. *IEEE Network*, 2020, vol. 34, no. 3, p. 134–142. DOI: 10.1109/MNET.001.1900287

[4] O'SHEA, T., HOYDIS, J. An introduction to deep learning for the physical layer. *IEEE Transactions on Cognitive Communications and Networking*, 2017, vol. 3, no. 4, p. 563–575. DOI: 10.1109/TCCN.2017.2758370

[5] HOSSEN, A., AL-WADAHI, F., JERVASE, J. A. Classification of modulation signals using statistical signal characterization and artificial neural networks. *Engineering Applications of Artificial Intelligence*, 2007, vol. 20, no. 4, p. 463–472. DOI: 10.1016/j.engappai.2006.08.004

[6] LE, T. T. *Applications of Deep Learning Framework and Localization to Intelligent Radio Spectrum Monitoring*. Chulalongkorn University, Master's Thesis, 2022. DOI: 10.58837/CHULA.THE.2022.135

[7] DAI, L., JIAO, R., ADACHI, F., et al. Deep learning for wireless communications: an emerging interdisciplinary paradigm. *IEEE Wireless Communications*, 2020, vol. 27, no. 4, p. 133–139. DOI: 10.1109/MWC.001.1900491

[8] O'SHEA, T. J., WEST, N. Radio machine learning dataset generation with GNU Radio. In *Proceedings of the 6th GNU Radio Conference*. 2016, p. 1–6. [Online] Cited 2026-08-24. Available at: https://pubs.gnuradio.org/index.php/grcon/article/view/11

[9] ZHANG, Z., LUO, H., WANG, C., et al. Automatic modulation classification using CNN-LSTM based dual-stream structure. *IEEE Transactions on Vehicular Technology*, 2020, vol. 69, no. 11, p. 13521–13531. DOI: 10.1109/TVT.2020.3030018

[10] WANG, N., LIU, Y., MA, L., et al. Multidimensional CNN-LSTM network for automatic modulation classification. *Electronics*, 2021, vol. 10, no. 14, p. 1–14. DOI: 10.3390/electronics10141649

[11] ZHOU, Q., JING, X., HE, Y., et al. LSTM-based automatic modulation classification. In *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*. Paris (France) – online, 2020, p. 1–4. DOI: 10.1109/BMSB49480.2020.9379677

[12] CHEN, Y., SHAO, W., LIU, J., et al. Automatic modulation classification scheme based on LSTM with random erasing and attention mechanism. *IEEE Access*, 2020, vol. 8, p. 154290–154300. DOI: 10.1109/ACCESS.2020.3017641

[13] DOSOVITSKIY, A., BEYER, L., KOLESNIKOV, A., et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv*, 2020, p. 1–32. DOI: 10.48550/arXiv.2010.11929

[14] ZHAO, C., CHEN, J., HUANG, X., et al. A cross-scale embedding based fusion transformer for automatic modulation recognition. *IEEE Communications Letters*, 2024, vol. 28, no. 1, p. 68–72. DOI: 10.1109/LCOMM.2023.3331265

[15] MA, W., CAI, Z., WANG, C. A transformer and convolution-based learning framework for automatic modulation classification. *IEEE Communications Letters*, 2024, vol. 28, no. 6, p. 1392–1396. DOI: 10.1109/LCOMM.2024.3380623

[16] ZHANG, X., LUO, Z., XIAO, W. CNN-BiLSTM-DNN-based modulation recognition algorithm at low SNR. *Applied Sciences*, 2024, vol. 14, no. 13, p. 1–14. DOI: 10.3390/app14135879

[17] RASHVAND, N., WITHAM, K., MALDONADO, G., et al. Enhancing automatic modulation recognition for IoT applications using transformers. *IoT*, 2024, vol. 5, no. 2, p. 212–226. DOI: 10.3390/iot5020011

[18] ZHANG, W., XUE, K., YAO, A., et al. CTRNet: An automatic modulation recognition based on transformer-CNN neural network. *Electronics*, 2024, vol. 13, no. 17, p. 1–25. DOI: 10.3390/electronics13173408

[19] DOBRE, O. A., ABDI, A., BAR-NESS, Y., et al. Survey of automatic modulation classification techniques: classical approaches and new trends. *IET Communications*, 2007, vol. 1, no. 2, p. 137–156. DOI: 10.1049/iet-com:20050176

[20] LIU, X., SONG, Y., ZHU, J., et al. An efficient deep learning model for automatic modulation classification. *Radioengineering*, 2024, vol. 33, no. 4, p. 713–720. DOI: 10.13164/re.2024.0713

[21] POLAK, L., TURAK, S., SOTNER, R., et al. Exploring deep learning architectures for RF signal classification. In *35th International Conference Radioelektronika (RADIOELEKTRONIKA)*. Hnanice (Czech Republic), 2025, p. 1–6. DOI: 10.1109/RADIOELEKTRONIKA64660.2025.11008396

[22] WU, S., WANG, C., LIANG, J., et al. A transformer-based framework with complex-valued convolution and enhanced Bi-LSTM for automatic modulation recognition. *Physical Communication*, 2025, vol. 76, p. 1–10. DOI: 10.1016/j.phycom.2025.102824

[23] EL-HARYQY, N., KHARBOUCHE, A., OUAMNA, H., et al. Improved automatic modulation recognition using deep learning with additive attention. *Results in Engineering*, 2025, vol. 26, p. 1–14. DOI: 10.1016/j.rineng.2025.104783

[24] ZHANG, L., HAO, X., LIU, Q., et al. Automatic modulation recognition of radio frequency proximity sensor signals based on adaptive relational graph attention network. *IEEE Transactions on Instrumentation and Measurement*, 2025, vol. 74, p. 1–13. DOI: 10.1109/TIM.2025.3552477

[25] HUO, Y., WANG, C., LIANG, J., et al. SMTrans: An efficient automatic modulation recognition network based on the scale-aware modulation transformer. *Physical Communication*, 2026, vol. 74, p. 1–12. DOI: 10.1016/j.phycom.2025.102966

[26] CHENG, R., JUAN, S., HUANG, M. L. TriFusion-Lite: A temporal-frequency-phase fusion lightweight network for modulation recognition. *Radioengineering*, 2026, vol. 35, no. 2, p. 304–314. DOI: 10.13164/re.2026.0304

[27] LIU, K., GAO, W., HUANG, Q. Automatic modulation recognition based on a DCN-BiLSTM network. *Sensors*, 2021, vol. 21, no. 5, p. 1–17. DOI: 10.3390/s21051577

[28] LIU, X., YANG, D., GAMAL, A. E. Deep neural network architectures for modulation classification. In *51st Asilomar Conference on Signals, Systems, and Computers*. Pacific Grove (CA, USA), 2017, p. 915–919. DOI: 10.1109/ACSSC.2017.8335483

[29] ZHANG, Y., ZHOU, Z., CAO, Y., et al. MAMC-optimal on accuracy and efficiency for automatic modulation classification with extended signal length. *IEEE Communications Letters*, 2024, vol. 28, no. 12, p. 2864–2868. DOI: 10.1109/LCOMM.2024.3474519

[30] KE, Z., VIKALO, H. Real-time radio technology and modulation classification via an LSTM auto-encoder. *IEEE Transactions on Wireless Communications*, 2021, vol. 21, no. 1, p. 370–382. DOI: 10.1109/TWC.2021.3095855

[31] XU, J., LUO, C., PARR, G., et al. A spatiotemporal multi-channel learning framework for automatic modulation recognition. *IEEE Wireless Communications Letters*, 2020, vol. 9, no. 10, p. 1629–1632. DOI: 10.1109/LWC.2020.2999453

[32] ZHANG, F., LUO, C., XU, J., et al. An efficient deep learning model for automatic modulation recognition based on parameter estimation and transformation. *IEEE Communications Letters*, 2021, vol. 25, no. 10, p. 3287–3290. DOI: 10.1109/LCOMM.2021.3102656

[33] ZHANG, J., WANG, T., FENG, Z., et al. AMC-Net: An effective network for automatic modulation classification. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Rhodes (Greece), 2023, p. 1–5. DOI: 10.1109/ICASSP49357.2023.10097070

[34] CHEN, Y., DONG, B., LIU, C., et al. Abandon locality: Frame-wise embedding aided transformer for automatic modulation recognition. *IEEE Communications Letters*, 2023, vol. 27, no. 1, p. 327–331. DOI: 10.1109/LCOMM.2022.3213523

About the Authors . . .

Sihem SOUIKI obtained her PhD in Telecommunications, from Abou Bekr Belkaid University, Tlemcen, Algeria, in 2015. Since 2018 she has been assistant professor of Communication Engineering and Telecommunication Networks in Belhadj Bouchaib university, Ain temouchent, Algeria. Member of STIC laboratory in the University of Tlemcen. Sihem SOUIKI is interested in the following research topics: Artificial Intelligence, cybersecurity, machine learning and deep learning, underwater wireless sensor networks, routing, energy optimization.

Mohammed Hicham HACHEMI graduated with an Engineering Telecommunications degree from the University of Saïda (Algeria) in 2007. Then he obtained a magister's degree in Electronics in 2011 at the Doctoral School of Information Science, Technology, and Telecommunications at the University of Sidi Bel Abbès. Mr. HACHEMI got his PhD in Telecommunications in 2017 from University of Tlemcen (UABT), Algeria. He is a lecturer and researcher at University of Science and Technology of Oran Mohamed Boudiaf (USTO- MB) since 2017, Algeria. Achieving his Habilitation to Direct Research (HDR) in 2022 allowed him to conduct many extensive research projects. His research domains include artificial intelligence, machine learning, deep learning technology, and research related to autonomous and connected vehicles. He also works on cognitive radio networks and wireless communications (5G/6G and beyond), as well as the design, optimization, and performance analysis of future antenna systems.

Mohammed MERZOUG received an engineering degree in Computer Science from Abou Bekr Belkaid university of Tlemcen, Algeria, in 2003, and M.S. degree in Information

and Communication Sciences in 2009, from the same University. After that, he joined the service oriented computing group in LRIT Laboratory, Tlemcen, Algeria, where he started his Ph.D.thesis. In Juin, 2018, he obtained the Ph.D. degree in Computer Science from University of Tlemcen, Algeria. His research interests include Artificial intelligence, deep learning, Image processing.

Mourad HADJILA received his Ph.D. degree in Telecommunications in 2014 at the University of Tlemcen, Algeria. Since 2002, he has worked as a lecturer and member of the STIC laboratory at the same University. He has been habilitated to supervise research since 2019 in Telecommunications. He was appointed to the rank of professor in January 2025. His research interests are in IoT, Intrusion Detection Systems, Artificial Intelligence, Machine Learning, Deep Learning, eXplainable Artificial Intelligence and Wireless Sensor Networks.

Mohammed M'HAMEDI received his Engineering degree in Telecommunications from Abou Bekr Belkaid University of Tlemcen, Algeria, in 2007. He went on to obtain a Magister's degree in Computer Science and Space Telecommunications in 2009, followed by a Ph.D. in Computer Science from the same university. His research interests include the Internet of Things (IoT), E-health, Artificial Intelligence, and Image Processing.

Lamia BENLALDJ received an engineering degree in Electronics from Abou Bekr Belkaid University of Tlemcen, Algeria, in 1997, and M.S. degree in Signals and Systems in 2000, from the same University. In december, 2025, she obtained the Ph.D. degree in Telecommunications at the University of Tlemcen (UABT), Algeria. She is a member of the STIC laboratory at the same University since 2005. His research interests include Artificial intelligence, Wireless Body Area Networks and Speech processing.