Mixed Signal Recognition Network Based on FD-MCNN and BiLSTM

Jiantao DONG [1], Yuanhan LIANG [1], Jie WANG [2], Zheng MA [1]*

[1] The Key Lab of Information Coding and Transmission, Southwest Jiaotong University, Chengdu, Sichuan, China
[2] CETC Key Laboratory of ElectroMagnetic Operation and Application, Chengdu, Sichuan, China

liudongmydear@163.com, 2023211098@my.swjtu.edu.cn*, zma@swjtu.edu.cn, wangjie16@cetc.com.cn

Submitted January 29, 2026 / Accepted April 29, 2026 / Online first May 19, 2026

Abstract. *Mixed-signal recognition in realistic wireless environments is challenging because weak signal components are often masked by stronger ones. To address this issue, this paper proposes a hierarchical recognition framework that combines multi-domain feature disentanglement with temporal dependency modeling. Specifically, the proposed network extracts complementary features from the time, frequency, modulation, and energy domains, enabling more robust representation of mixed signals under complex interference conditions. Based on these features, the framework first identifies the dominant signal component, then enhances the weak component to reduce the masking effect, and finally employs a bidirectional long short-term memory network (BiLSTM) network for temporal modeling and classification. Experiments with signal-to-noise ratio (SNR) ranging from 0 to 30 dB show that the proposed method can effectively recognize both strong and weak signal components while improving overall robustness. These results demonstrate the effectiveness of the proposed framework for mixed-signal recognition in interference-rich wireless environments.*

Keywords

Mixed signal recognition, hierarchical recognition framework, deep learning, feature disentanglement

1. Introduction

In modern wireless environments, signals with different modulation formats often coexist and interfere within the same time–frequency resources [1–3]. Reliable modulation recognition is essential for spectrum awareness, signal decoding, and adaptive receiver design [4], [5]. With the increasing complexity of wireless systems, mixed-signal scenarios have become increasingly common. However, recognizing both dominant and weak components in mixed-signal scenarios remains challenging, particularly under time–frequency overlap and low SNR conditions.

Prior work has explored signal recognition from multiple perspectives, primarily relying on model-driven methods. For example, Hajek et al. proposed a new time-domain method for converting real-valued sinusoidal signals into analytic signals and, on this basis, developed a high-precision frequency estimation algorithm [6]. Even with a small number of samples, this frequency estimation algorithm can effectively compensate for various mixed errors unrecognized in traditional discrete Fourier transform (DFT); however, this algorithm is mainly designed for sinusoidal signals, and the sidelobe suppression ability of the rectangular window adopted in this algorithm is poor in the presence of strong interference components.

Nevertheless, such traditional methods are often constrained by strong prior assumptions and limited adaptability when dealing with complex modulation patterns and non-stationary mixed signals. In recent years, with the rapid development of data-driven techniques, deep learning-based methods have emerged as a promising alternative for signal recognition tasks, with studies focusing on either single-signal recognition or semi-supervised learning for signal processing.

Liu et al. developed an automatic modulation recognition method that balances high spatial resolution with low computational cost by introducing a multi-scale dilated convolution pyramid module (MDPM) combined with lightweight neural network techniques [7]. Experimental results show that this model performs exceptionally well on a complex signal dataset containing 18 modulation formats, achieving nearly 100% recognition accuracy at a SNR of 0 dB, with only 48,952 parameters. The drawback is that this study mainly focuses on modulation classification for single signal sources, and its performance in scenarios with multiple coexisting signals or severe co-channel interference in complex mixed signals still needs further verification.

For RF signal classification, Polak et al. tested CNN, GRU, and convolutional gated recurrent unit deep neural network (CGDNN) on a self-built MATLAB dataset [8]. CGDNN outperformed the others in most cases, while CNN excelled at Rician fading signals. All models achieved high accuracy for RF signals with basic impairments at moderate-to-high SNR, and the fused CGDNN showed robust feature extraction for sequential and spatial RF signal characteristics. However, all models performed

DOI: 10.13164/re.2026.0256

poorly on Rayleigh fading and mixed RF signals, with CNN being ineffective for Rayleigh fading and GRU demanding heavy computational resources and risking overfitting.

For scenarios with limited labeled samples, Pan et al. proposed a semi-supervised learning method based on generative adversarial networks, namely MIML-GAN (multi-instance multi-label generative adversarial network) [9]. This method combines the advantages of multi-instance multi-label learning and GANs, enhancing feature representation through adversarial learning on unlabeled samples, but it has not been validated for mixed-signal recognition with energy imbalance.

More relevant to this study, while the previous section has briefly mentioned the poor performance of deep learning methods in mixed-signal scenarios, several recent works focusing on mixed-signal processing and recognition still have obvious limitations in handling energy-imbalanced scenarios.

Focusing on mixed-signal separation, Yang et al. proposed a hybrid modulation recognition method based on cyclic spectrum projection and deep neural networks [10]. Although this method has good robustness to changes in signal symbol rate and mixed signal energy ratio (with an average recognition rate exceeding 95% when SNR $\geq$ 0 dB), its high computational complexity limits its practical application. Li et al. proposed a method based on improved separation matrix normalization [11], which enhances the mixed-signal separation effect by constructing a pseudo multi-channel structure with memory. Similarly, Yang et al. proposed a blind separation method for mixed signals combining multi-resolution singular spectrum analysis with improved principal component analysis [12], which demonstrates better robustness in unknown signal processing but does not address the recognition of weak components in energy-imbalanced scenarios.

Moving to mixed-signal recognition, Liu et al. proposed a mixed-signal recognition method based on a long short-term memory deep residual shrinkage network (LDRSN) [13]. This method integrated residual modules, shrinkage modules, and long short-term memory (LSTM) modules to construct a recognition framework that balances feature selection and temporal modeling capabilities. Experimental results showed that the proposed method achieved an average recognition rate of 92.7% on a dataset at 16 dB SNR, with 12 out of 21 classes of mixed signals achieving recognition rates close to 100%. On a real-world measured dataset, the recognition accuracy further reached 95.53%. However, this method did not specifically address the recognition of weak components in energy-imbalanced mixed signals.

Kang et al. proposed TMCCNN (two-component mixed-signal modulation classification CNN), one of the early representative works applying deep learning to mixed-signal recognition [14]. This method directly took the I/Q components of the signal as input, avoiding complex manual feature design and thereby preserving the original modulation characteristics more completely. Experiments showed that TMCCNN demonstrated strong robustness under low SNR conditions, achieving a recognition accuracy of 95.62% for dual-component signals at 0 dB. Nevertheless, its performance degrades significantly when the energy ratio between mixed components is highly imbalanced.

Xu and Lin proposed a deep learning-based method for mixed-signal modulation classification [15], focusing on energy-imbalanced mixed signals in non-orthogonal multiple access (NOMA) scenarios. They first reproduced a CNN model designed for single-signal modulation classification and then optimized its structure to extend it to mixed-signal recognition. Experimental results showed that the proposed method outperformed the single CNN model in low-to-medium SNR regions. However, the performance gain diminished with increasing training samples, and it still struggled to effectively recognize weak components under severe time-frequency overlap.

Despite recent progress in mixed-signal separation and recognition, two key limitations remain in existing methods. First, existing methods are prone to missing weak signal components when there is an energy imbalance between strong and weak components, especially under time–frequency overlap and low SNR conditions. Second, most methods either focus on mixed-signal separation but ignore weak component recognition, or fail to effectively fuse multi-domain features (time, frequency, modulation, and energy), leading to insufficient feature representation and difficulty in balancing the recognition performance of both strong and weak components.

To address these issues, we propose a hierarchical recognition framework, namely feature-disentangled multi-column convolutional neural network-bidirectional long short-term memory (FD-MCNN-BiLSTM), for mixed signals. The core innovations of FD-MCNN-BiLSTM are twofold: (1) The FD-MCNN module extracts complementary multi-domain features from time, frequency, modulation, and energy information, thereby realizing feature disentanglement between strong and weak components; (2) The BiLSTM module models temporal dependencies of the extracted features, thereby improving classification accuracy. Specifically, the framework first identifies the dominant component and then enhances the weak component through feature disentanglement and signal enhancement, thereby addressing the problem of weak component missing. Experimental results demonstrate that the proposed method achieves an average recognition accuracy improvement of 5.2%–8.7% for weak components and 2.1%–3.5% for dominant components across 0–30 dB SNR conditions, confirming its effectiveness in interference-rich and energy-imbalanced environments. To promoter producibility, the dataset and the source code for this simulation-based study are publicly available [16].

The remainder of this paper is organized as follows. Section 2 reviews the theoretical foundations of mixed-

signal recognition. Section 3 presents the detailed architecture and design principles of the proposed FD-MCNN-BiLSTM network. Section 4 describes the experimental setup, presents comparative results with state-of-the-art methods, and conducts an ablation study to validate the effectiveness of each component. Finally, Section 5 concludes the paper.

2. Research Foundation

2.1 Signal Mathematical Model

In wireless communication systems, modulation plays a fundamental role for reliable information transmission. It converts baseband signals into waveforms suitable for transmission over wireless channels by modifying specific properties of the carrier. In this paper, we use I/Q sequences as the basic signal representation, and their discrete complex baseband representation is given as [17]

$$x(n) = I(n) + jQ(n), n = 0,1,...,N-1. \quad (1)$$

The input signal has dimensions of $2 \times N$, where 2 represents the in-phase and quadrature braches, and N represents the length of the signal segment. The sample length in this paper is set to $N = 1024$, mainly for the following reasons. First, considering that the fast Fourier transform (FFT) is the basis for frequency-domain feature extraction in the FD-MCNN module, choosing 1024, a power of 2, facilitates efficient FFT computation. Second, this choice maintains a reasonable balance between information retention and computational complexity; specifically, shorter sequences usually fail to capture the necessary temporal context, while longer sequences significantly increase the computational burden of the model without a proportionate improvement in recognition performance [12], [13].

According to the variation pattern of signal parameters, modulation techniques can be broadly classified into digital and analog modulation. Digital modulation maps information onto discrete signal states through symbol-based representation. Representative digital modulation schemes include phase-shift keying (PSK), quadrature amplitude modulation (QAM), frequency-shift keying (FSK), and amplitude-shift keying (ASK). In contrast, analog modulation varies signal parameters continuously, with common schemes including amplitude modulation (AM), frequency modulation (FM), and phase modulation (PM).

The modulation set considered in this paper comprises AM, FM, 2ASK, 4FSK, and 16QAM. For convenience, let the carrier frequency be denoted as f_c, the sampling frequency as f_s, and the sampling interval as $T = 1/f_s$. The mathematical expressions for these five types of modulated signals are given by [18]

$$\begin{aligned}
x_{\mathrm{AM}}(t) &= \left[A_0 + m(t)\right]\cos\left(2\pi f_c t + \phi\right), \\
x_{\mathrm{FM}}(t) &= A\cos\left[2\pi f_c t + 2\pi k_f \int_0^t m(\tau)\mathrm{d}\tau + \phi\right], \\
x_{\mathrm{ASK}}(t) &= \left[A\sum_k a_k g(t - kT_s)\right]\cos(2\pi f_c t + \phi), \\
x_{\mathrm{FSK}}(t) &= A\cos\left[2\pi\left(f_c + \Delta f_{a_k}\right)t + \phi\right], \\
x_{\mathrm{QAM}}(t) &= I\cos(2\pi f_c t) - Q\sin(2\pi f_c t)
\end{aligned}, \quad (2)$$

where A_0 denotes the DC component, $m(t)$ represents the baseband signal, Φ is the phase, k_f stands for the frequency modulation sensitivity, a_k indicates the amplitude values of the k-th symbol, and Δf_{a_k} represents symbol-dependent offset.

Mixed signals are composite waveforms formed by the superposition of multiple modulation components that arrive at the receiver within the same frequency band and time window.

In practical communication scenarios, channel attenuation, multipath propagation, and carrier frequency offsets often cause the received waveform to contain multiple overlapping components, resulting in mixed signals. Let the received mixed signal be [10]

$$y(n) = \sum_{m=0}^{M-1} \alpha_m x_m(n - \tau_m) \mathrm{e}^{\mathrm{j}(2\pi\Delta f_m nT + \phi_m)} + w(n) \quad (3)$$

where $x_m(n)$ denotes the baseband signal of the m-th modulation component, α_m is the amplitude of the m-th signal, τ_m is the time delay, f_m is the frequency offset, ϕ_m is the phase offset, and $w(n)$ is the additive noise.

In the mixed signal model, the time-domain, frequency-domain, and modulation-specific characteristics of different components can become tightly coupled, making it difficult for traditional time-domain or frequency-domain feature extraction methods to obtain sufficiently discriminative information.

Moreover, because the constituent signals often exhibit different power levels, the dominant component may drive the overall recognition outcome, while weaker components are easily masked by the stronger ones.

2.2 Time-Frequency Analysis Method for Signals

Time-frequency analysis is an important tool in signal processing, as it provides a joint representation of signals in the time and frequency domains. Conventional time-domain and frequency-domain methods have inherent limitations, and neither of them can effectively characterize the time-varying spectral properties of nonstationary signals. This limitation is particularly evident for complex mixed signals, in which multiple modulated components may overlap within the same time interval and frequency band, making accurate discrimination difficult for traditional methods.

By mapping a signal into a two-dimensional time-frequency representation, time-frequency analysis can simultaneously reveal its instantaneous frequency, energy distribution, and local structural characteristics. This richer representation provides more informative features for the accurate identification of mixed signals.

The short-time Fourier transform (STFT) is one of the most widely used methods in time-frequency analysis. It extends the conventional Fourier transform (FT) by introducing a sliding window to characterize how the spectral content of a nonstationary signal evolves over time. Specifically, localized FT are computed within successive windowed segments, thereby yielding a time-varying spectral representation. The discrete form of the STFT is given by [19]

$$X(m,k)=\sum_{n=0}^{L-1}x(n+mH)w(n)\,\mathrm{e}^{-\mathrm{j}2\pi kn/L} \tag{4}$$

where L denotes the window length, H denotes the frame shift, m denotes the time-frame index, and k denotes the frequency index. The window function $w(n)$ is typically selected from functions such as the Hamming window or the Hann window; in this paper, the Hamming window is adopted. The window length L and frame shift H should be determined according to the temporal and spectral characteristics of the signal.

Applying the STFT yields a time-frequency spectrogram that describes how the spectral content of a signal evolves over time. In the spectrogram, the horizontal axis denotes time, the vertical axis denotes frequency, and the color intensity reflects the signal energy at the corresponding time-frequency coordinates. The time-frequency resolution of the STFT is determined mainly by the window length and frame shift; therefore, the selection of these parameters involves a trade-off between temporal resolution and frequency resolution.

The spectrogram of mixed signals reveals the overlap of multiple components within the same time-frequency region. By examining the time-frequency characteristics of mixed signals, several prominent features can be identified:

(1) Strong-signal masking effect: When high-energy components dominate, the time–frequency features of weaker signals can be masked and become difficult to observe. This effect is pronounced at low SNR, motivating more refined feature extraction methods to recover weak-signal information from the complex background.

(2) Local overlap characteristics: Different modulated signals may partially overlap in the time–frequency domain, forming complex spectral textures. Components can share portions of the frequency band or overlap within the same time interval, which makes conventional global frequency-domain analysis ineffective for distinguishing them.

(3) Separability of signal characteristics: Although mixed signals overlap in the time–frequency domain, different modulation schemes often preserve distinctive spectral signatures. For example, AM signals exhibit characteristic bandwidth patterns, whereas FM signals exhibit continuously varying instantaneous frequency. Such differences provide useful cues for deep learning-based multi-column feature learning.

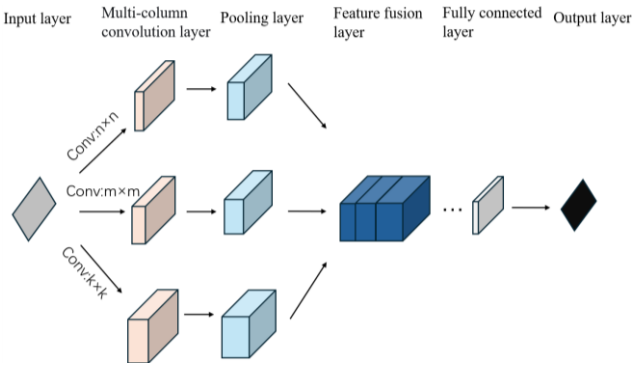


Fig. 1. Architecture diagram of the MCNN.

2.3 MCNN-Based Deep Learning Model

The MCNN extends the conventional CNN by employing multiple parallel branches for feature extraction. Each branch uses convolutional kernels with different sizes or receptive fields to capture complementary information at multiple scales [20], [21]. This architecture improves the network's ability to learn discriminative representations from complex signals exhibiting multi-scale characteristics. As shown in Fig. 1, MCNN extracts features in parallel through multiple branches, with each branch performing a distinct convolutional operation. The resulting features are then concatenated to form a multi-scale feature representation. This design enables the network to capture both local details and broader structural information within a unified framework, making it well suited for signal recognition tasks involving diverse spectral characteristics.

Based on the above signal models, time-frequency analysis method, and MCNN architecture, the FD-MCNN-BiLSTM hierarchical recognition framework is developed in the following section.

3. Proposed FD-MCNN-BiLSTM Framework

To address the challenge of recognizing highly overlapping mixed signals in the time-frequency domain, this paper proposes a hierarchical FD-MCNN-BiLSTM recognition framework. As shown in Fig. 2, the proposed framework consists of four main modules: data acquisition and preprocessing, the primary recognition network, the signal enhancement module, and the secondary recognition network.

The model adopts a two-stage strategy: it first classifies high-energy components using FD-MCNN with cross-branch attention, and then identifies low-energy components from the residual signal using a BiLSTM, enabling dual-signal recognition. The I/Q time-domain sequences

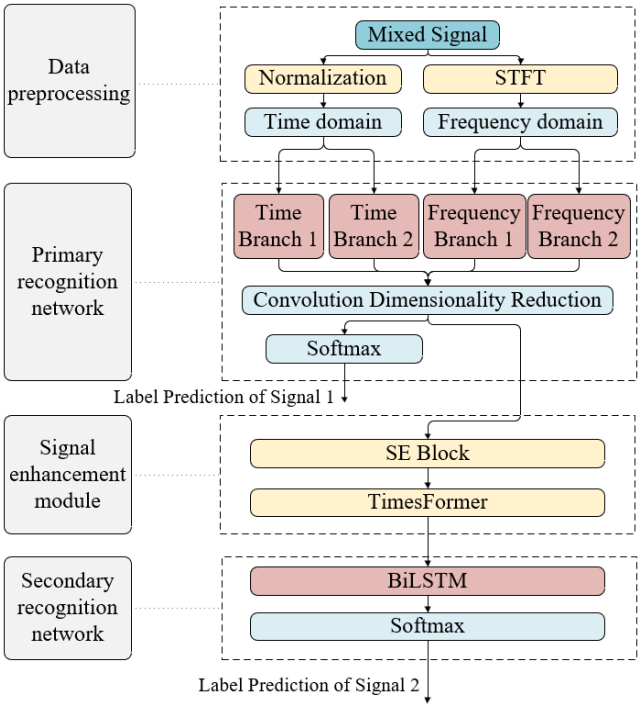


Fig. 2. Overall architecture of the proposed framework for mixed signals with different energy levels.

exhibit distinctive, correlated patterns that provide discriminative cues for modulation identification.

3.1 Data Preprocessing

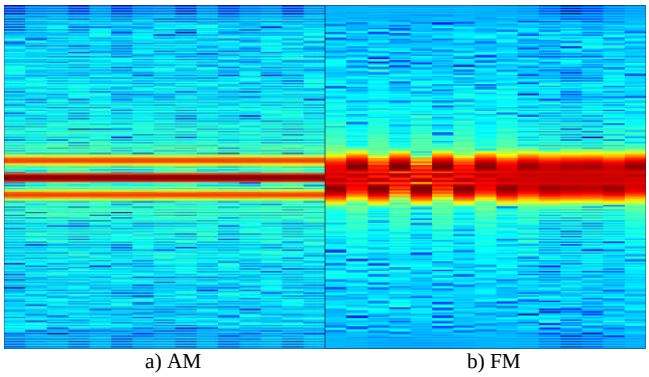
The proposed framework preprocesses each mixed-signal sample to generate suitable input representations for hierarchical recognition. Since mixed signals are highly overlapping and nonstationary, both one-dimensional time-domain sequences and two-dimensional time–frequency representations are constructed to provide complementary information.

First, the received signal is represented as an I/Q sequence, which preserves the original temporal structure and the correlation between the in-phase and quadrature components. These one-dimensional sequences are directly fed into the network to enable automatic temporal feature learning without relying on handcrafted features.

To reduce amplitude scale variations and improve training stability, min-max normalization is applied to the input signal, as given by [22]

$$x_i' = \frac{x_i - \min(x(n))}{\max(x(n)) - \min(x(n))}, \quad i = 0,1,\ldots,L-1. \quad (5)$$

In addition to the normalized one-dimensional time-domain input, the proposed framework transforms the original signals into two-dimensional time–frequency representations. This is because mixed signals often exhibit severe overlap in the original domain, making it difficult to distinguish different modulation components using only temporal or spectral information.



a) AM b) FM

Fig. 3. Time-frequency spectrograms of AM and FM signals.

By projecting the signals into the time-frequency domain, the local energy distribution and modulation-dependent spectral structures become more distinguishable, which is beneficial for subsequent feature extraction and classification.

In this paper, STFT is adopted to generate the time–frequency representation. By applying localized FT with a sliding window, STFT provides an effective joint description of the temporal and spectral characteristics of nonstationary signals. The resulting spectrograms highlight the differences among modulation components and provide informative two-dimensional inputs for the FD-MCNN.

Taking AM and FM signals as examples, the corresponding time-frequency spectrograms are shown in Fig. 3. Compared with raw waveforms, these spectrograms reveal clearer modulation-specific patterns, thereby improving the separability of different signal components and supporting more accurate hierarchical recognition.

After preprocessing, each mixed-signal sample is represented by a normalized one-dimensional I/Q sequence and a two-dimensional time-frequency spectrogram, which are jointly used as inputs to the proposed framework.

3.2 Primary Recognition Network

Mixed signals exhibit strong feature coupling because multiple modulation components are superposed in both the time and frequency domains. To better extract and fuse such heterogeneous features, an MCNN-based architecture is adopted as the backbone. Compared with single-stream backbones such as ResNet or dense convolutional network (DenseNet), the multi-branch design of MCNN is better suited to jointly modeling temporal, spectral, modulation-related, and energy-related characteristics.

To address modulation coupling, energy imbalance, and feature heterogeneity, we propose an improved multi-branch convolutional network termed FD-MCNN. The network structure is shown in Fig. 4.

The network is designed to jointly exploit temporal, spectral, modulation-related, and energy-related information through domain-informed specialized branches and cross-branch interactions, thereby enhancing feature com-

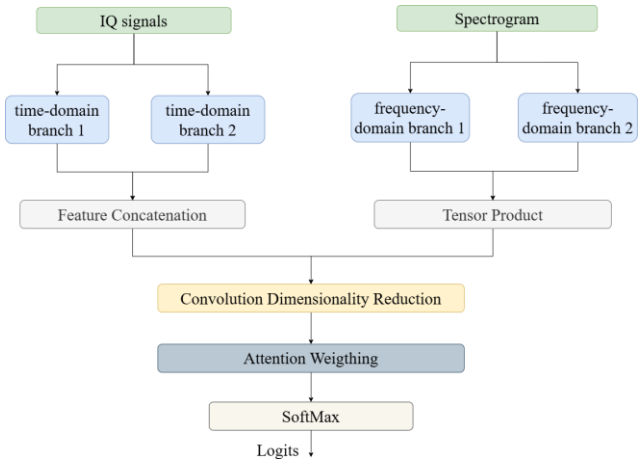


Fig. 4. FD-MCNN architecture diagram.

plementarity and improving generalization in complex overlapping scenarios.

As illustrated in Fig. 4, FD-MCNN adopts a dual-path input structure, in which the one-dimensional I/Q sequence and the corresponding STFT-based time–frequency representation are processed in parallel.

Based on these two inputs, the network comprises four specialized branches: (1) a time-domain branch for capturing local transient dynamics in I/Q signals; (2) a frequency-domain branch for extracting global structural information from time–frequency representations, such as energy concentration and frequency drift; (3) a modulation-characteristics branch for learning modulation-sensitive patterns and reducing ambiguity among mixed modulation components; and (4) an energy-sensing branch for enhancing robustness under low-SNR conditions by modeling local energy distributions and dynamic-range variations.

Through the joint action of these branches, FD-MCNN produces a discriminative high-dimensional representation that effectively separates coupled modulation and energy characteristics in overlapping environments, thereby enabling more accurate first-stage recognition of high-energy signal components.

The one-dimensional and two-dimensional inputs are processed by four dedicated branches and then fused to form the final representation. The detailed designs of these branches are described as follows.

(1) Time-Domain Branch Design

Table 1 summarizes the architecture of the time-domain branch. In Tab. 1, $n \times a$ denotes the number of convolutional kernels (n) and the kernel size (a). ReLU is used in all convolutional layers and in the first two fully connected layers.

Guided by the Bedrosian theorem [23], the temporal branch uses large, residual, and dilated convolutions to expand the receptive field 2–3 times the signal period, capturing slow envelope variations, localizing transients, and modeling large-scale temporal trends, effectively extracting envelope and amplitude features while addressing energy dominance.

Network layer	Parameter($n \times a$)	Activation Function
Conv1D + BN	16 × 7	ReLu
Conv1D + BN	32 × 5	ReLu
Conv1D + BN	64 × 3	ReLu

Tab. 1. Time-domain branch diagram.

Network layer	Parameter($n \times a \times b$)	Activation Function
Conv2D + BN	16 × 3 × 1	ReLu
Conv2D + BN	32 × 1 × 3	ReLu
Max Pooling	2 × 2	-

Tab. 2. Frequency-domain branch diagram.

(2) Frequency-Domain Branch Design

The architecture of the frequency-domain branch is summarized in Tab. 2. In Tab. 2, $n \times a \times b$ denotes the number and size of convolutional kernels, while 2×2 specifies max-pooling window dimensions. ReLU is applied throughout convolutional and fully connected layers.

Based on the time–frequency equivalence principle from Plancherel's theorem [24], the time-domain signal is converted to a time-frequency representation using STFT with a Hamming window ($N = 256$), retaining phase information. A 2D convolutional network then extracts energy distribution patterns from the spectrogram to address the discrimination of frequency-modulated signals.

The convolutional module employs asymmetric kernels: vertical (3×1) kernels strengthen frequency-band correlations, and horizontal (1×3) kernels capture temporal dynamics. This design supports demodulation of continuous phase variations and detection of discrete frequency-hopping features.

(3) Modulation-Characteristics Branch Design

The architecture of the modulation-characteristics branch is detailed in Tab. 3. In Tab. 3, $n \times a$ denotes the number and size of convolutional kernels, and Channel Attention specifies the reduction ratio. ReLU is used in all convolutional layers and the first two fully connected layers.

(4) Energy-Sensing Branch Design

The architecture of the energy-sensing branch is presented in Tab. 4.

Network layer	Parameter ($n \times a$)	Activation Function
SeparableConv1D + BN	16 × 5	ReLu
SeparableConv1D + BN	32 × 3	ReLu
Channel Attention	8	-

Tab. 3. Modulation-characteristics branch diagram.

Network layer	Parameter($n \times a \times b$)	Activation Function
Conv2D + BN	16 × 5 × 1	ReLu
Conv2D + BN	32 × 1 × 5	ReLu
Max Pooling	2 × 2	-

Tab. 4. Energy-sensing branch diagram.

Leveraging Parseval's theorem [25], an energy-normalization pathway is built through local energy statistics and dynamic range compression to enhance low-SNR discrimination. In Tab. 4, $n \times a \times b$ represents the number and size of kernels, and 2×2 denotes max-pooling windows; ReLU is applied in both convolutional and fully connected layers.

After branch-wise feature extraction, the resulting representations are integrated through a dual-path fusion strategy. Specifically, the time-domain and modulation-related features are concatenated to preserve their complementary physical characteristics, whereas the frequency-domain and energy-related features are combined through tensor-product operations to capture nonlinear correlations. The fused representation is then refined by convolutional dimensionality reduction and attention-based weighting to adapt to different energy ratios.

Finally, a Softmax classifier outputs the predicted category of the high-energy component, thereby completing the first-stage recognition of mixed signals.

3.3 Signal Enhancement Module

After the primary recognition network extracts discriminative features of the dominant signal component, weak-signal-related information may still be partially suppressed due to energy imbalance and feature coupling in overlapping mixtures.

To alleviate this problem, a Channel-Spatiotemporal Attention (CSTA) module is introduced, as illustrated in Fig. 5. Positioned between FD-MCNN-based feature extraction and the second-stage recognition network, the CSTA module jointly models channel-wise and spatiotemporal dependencies to selectively enhance weak-signal representations and improve their distinguishability in complex overlapping environments.

Specifically, the proposed CSTA module integrates an SEBlock and a TimeSformer-based attention unit [26], [27]. The SEBlock focuses on channel-wise feature recalibration, whereas the TimeSformer-based unit captures long-range dependencies in the spatiotemporal feature space. Through the combination of these two mechanisms, the module can effectively suppress redundant interference information while strengthening weak but task-relevant features.

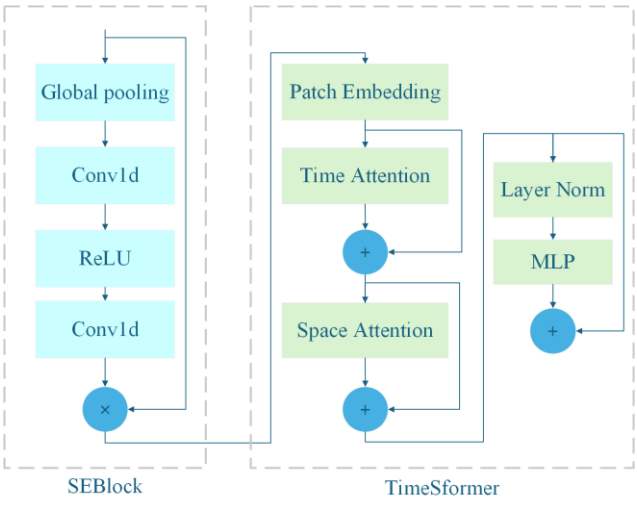


Fig. 5. Architecture of the signal enhancement module.

The detailed designs of these two attention components (SEBlock and TimeSformer) and the output of the enhancement module are described as follows.

(1) Channel Attention (SEBlock)

The channel attention branch is implemented using an SEBlock. By applying global pooling followed by channel transformation operations, the branch generates adaptive channel-wise weights for the intermediate feature maps. In this way, channels that are more relevant to weak-signal characteristics are assigned larger weights, while less informative or interference-dominated channels are suppressed. This channel recalibration mechanism helps the network focus on subtle but discriminative features that may otherwise be overwhelmed by dominant signal components.

(2) Spatiotemporal Attention (TimeSformer-based)

To further enhance the representation of weak signals, a TimeSformer-based spatiotemporal attention mechanism is introduced. The intermediate features are reshaped into token sequences and processed by self-attention operations along the temporal and structural dimensions. Unlike local convolution, self-attention can capture long-range dependencies across different positions, thereby enabling the model to identify weak but correlated patterns distributed over the feature sequence. This mechanism is particularly beneficial when weak-signal features are partially masked by strong-signal interference or background noise.

(3) Output of the Enhancement Module

After channel attention and spatiotemporal attention are applied, the CSTA module outputs an enhanced feature representation in which weak-signal-related information is selectively emphasized and interference is further suppressed. This refined representation serves as the input to the subsequent secondary recognition network, providing a more informative basis for weak-signal identification.

3.4 Secondary Recognition Network

To further model temporal dependencies in the feature representations enhanced by the CSTA module, a BiLSTM is introduced as the secondary recognition network, as illustrated in Fig. 6. As the second-stage classifier, the BiLSTM takes the sequential features output by the CSTA module as input and captures long-range contextual information that is beneficial for the recognition of weak signal components in overlapping scenarios.

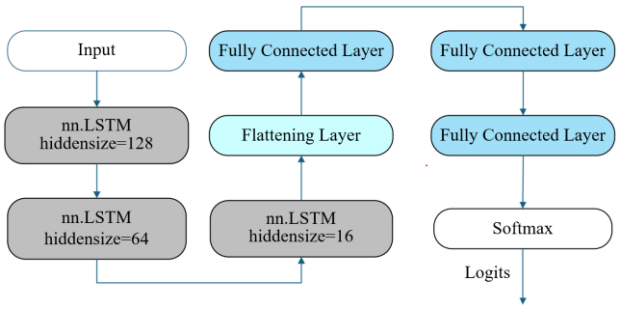


Fig. 6. BiLSTM architecture diagram.

Network layer	Parameter(*n*)	Activation Function
BiLSTM	128	-
BiLSTM	64	-
BiLSTM	16	-
Flattening	-	-
Fully connected	64	ReLu
Fully connected	16	ReLu
Fully connected	3	Softmax

Tab. 5. BiLSTM structure.

A BiLSTM consists of forward and backward LSTM units, whose hidden states are concatenated at each time step to exploit both past and future contextual information [28–30]. This bidirectional modeling strategy is particularly suitable for mixed signal recognition, where the temporal evolution of weak but discriminative features should be preserved despite interference from dominant components. By integrating information from both directions, the BiLSTM generates feature sequences enriched with more complete temporal dependencies, which greatly facilitates the subsequent identification of weak modulation components.

After bidirectional temporal modeling, the output features are passed through a fully connected layer and a Softmax classifier to generate the final recognition results. The gating mechanism of the BiLSTM helps mitigate the influence of noise and preserve discriminative temporal dependencies, thereby improving the robustness of the network under low-SNR conditions.

Table 5 lists the structural parameters of the BiLSTM network. In Tab. 5, *n* denotes hidden layers. The intermediate fully connected layer uses ReLU, where the output layer uses Softmax.

4. Experimental Results and Analysis

4.1 Dataset Construction

The dataset used in this paper was collected from real-world wireless environments using a Universal Software Radio Peripheral – USRP-based acquisition platform. As shown in Fig. 7, two transmitters simultaneously emitted signals with different modulation types and unequal energy levels, and the resulting mixed signals were received and recorded by the USRP receiver. This acquisition

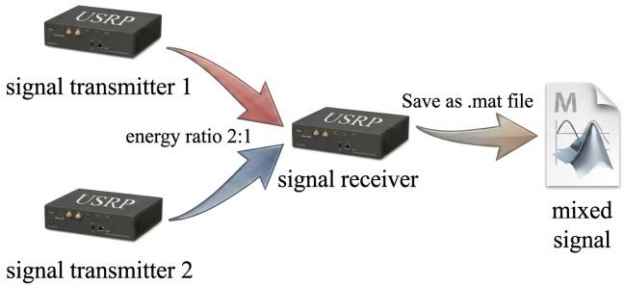


Fig. 7. Signal acquisition.

setup was designed to simulate realistic mixed-signal scenarios in which multiple emitters coexist and their signal components overlap in both time and frequency.

The indoor measurement environment is shown in Fig. 8, and a close-up view of the NI USRP-2922 device used in the experiments is presented in Fig. 9. The measurement campaign was conducted in an indoor room of approximately 150 m². The propagation condition was mainly line-of-sight, while reflections from surrounding objects introduced mild multipath effects. During data acquisition, the two transmitters and the receiver were placed at approximately the same height, and the distance between each transmitter and the receiver was about 1.2 m.

All experiments were performed on an NI USRP-2922 platform equipped with external antennas. The transmitted waveforms were generated in NI LabVIEW

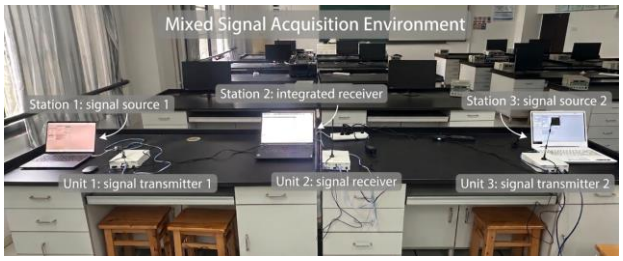


Fig. 8. Photograph of the experimental environment.

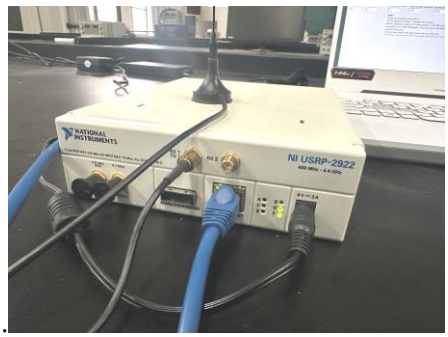


Fig. 9. Close-up view of the NI USRP-2922 device used in the experiments.

	Carrier Frequency	Sampling Rate	TX-Gain	RX-Gain
Transmitter 1	2450 MHz	500 kHz	7 dB	-
Transmitter 2	2450 MHz	500 kHz	10 dB	-
Receiver	2450 MHz	500 kHz	-	0 dB

Tab. 6. Transmitter and receiver parameter table.

using the built-in USRP modulation examples. For each transmitter, the baseband signal was digitally generated in software and configured according to the predefined modulation settings described earlier. The generated baseband waveforms were then upconverted and transmitted simultaneously by two USRP devices, so that mixed signals were formed through over-the-air superposition and captured by the receiver. The received I/Q data were stored as .mat files for subsequent dataset construction. The main configuration parameters of the transmitters and receiver are listed in Tab. 6.

The carrier frequency of both transmitters and the receiver was set to 2450 MHz, and the sampling rate was 500 kHz. Since the LabVIEW demonstration program did not explicitly provide a parameter for directly setting the absolute transmit power, the relative power of the two transmitted signals was controlled by adjusting the transmitter (TX) gains of the two transmitters. Specifically, the TX gains of Transmitter 1 and Transmitter 2 were set to 7 dB and 10 dB, respectively. In addition, the gain of the receiver (RX) was fixed at 0 dB. The TX gains difference of 3 dB corresponds approximately to a 2 : 1 power ratio, which was used to construct mixed signals with unequal component strengths.

Based on the above measurement setup, mixed-signal data were collected for the modulation combinations considered in this study, resulting in 20 mixed-signal categories under the specified relative power setting. To evaluate the proposed method under different noise conditions, additional controllable noise was introduced only at the receiver side during software post-processing after signal acquisition, and no artificial noise was added at the transmitter side. Specifically, artificial noise was superimposed on the received mixed signal to obtain seven SNR levels: 0, 5, 10, 15, 20, 25, and 30 dB. In this study, the SNR is defined as the ratio of the power of the received mixed signal, i.e., the superposed signal from TX1 and TX2 at the receiver, to the power of the artificially added noise. Therefore, the reported SNR values are specified for the overall received mixed signal rather than for either transmitter individually.

After acquisition, each collected modulation dataset was saved as a .mat file, with each dataset containing 5,000,000 complex sampling points. The dataset was partitioned into individual samples, with each sample containing 1024 sampling points. These samples were then divided into training, validation, and tested subsets at a ratio of 6 : 2 : 2 for the follow-up experiments. Each sample was then processed according to the preprocessing procedure described in Sec. 3.1, including normalization and STFT. These processed data were used as the one-dimensional and two-dimensional inputs to the proposed hierarchical recognition framework.

4.2 Evaluation Metrics for Mixed-Signal Recognition

In the mixed signal recognition task considered in this study, each sample contains two modulation labels. Therefore, conventional single-label accuracy is not sufficient to fully reflect the recognition performance. To fairly evaluate the model under this dual-label setting, a customized accuracy metric is adopted, defined as

$$P_{\mathrm{r}} = \frac{\mathrm{NUM}_{\mathrm{two}} + 0.5 \times \mathrm{NUM}_{\mathrm{one}}}{\mathrm{NUM}_{\mathrm{sum}}} \times 100\% \tag{6}$$

where $\mathrm{NUM}_{\mathrm{two}}$ denotes the number of test samples for which both modulation types are correctly recognized, $\mathrm{NUM}_{\mathrm{one}}$ denotes the number of test samples for which only one modulation type is correctly recognized, and $\mathrm{NUM}_{\mathrm{sum}}$ is the total number of test samples.

In addition to accuracy, the F1-score is employed to provide a more comprehensive evaluation of the recognition performance. The F1-score is the harmonic mean of Precision and Recall and reflects the balance between false positives and false negatives. F1-score is defined as [31]

$$\mathrm{F1} = \frac{2 \times \mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \tag{7}$$

where Precision and Recall are defined as TP/(TP + FP) and TP/(TP + FN), respectively, and TP, FP, and FN denote the numbers of true positives, false positives, and false negatives.

4.3 Experimental Results and Analysis

During training, a multi-label binary cross-entropy loss is employed to handle samples containing two modulation labels. This loss computes the cross-entropy for each label independently and then averages them, thereby enabling the network to optimize the prediction of each modulation component, which is defined as [32]

$$\mathrm{Loss} = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot \log\left(p_i\right) + \left(1 - y_i\right)\log\left(1 - p_i\right). \tag{8}$$

The Adam (Adaptive Moment Estimation) optimizer is adopted to update the network parameters using adaptive learning rates and first- and second-order moment estimates of the gradients, which facilitates efficient and stable training. In this study, mixed signal datasets at SNR levels of 0, 5, 10, 15, 20, 25, and 30 dB are used to train and evaluate the proposed FD-MCNN-BiLSTM as well as three representative comparison models, namely LDRSN [13], TMCCNN [14], and CNN-LSTM [15]. The network configurations of these models are summarized in Tab. 7.

All experiments in this paper were conducted under the same experimental conditions, as detailed in Tab. 8.

Figure 10 shows the loss and accuracy curves of FD-MCNN-BiLSTM, LDRSN, TMCCNN, and CNN-LSTM under 30 dB.

Overall, proposed FD-MCNN-BiLSTM exhibits a faster and more stable decrease in loss during the early training epochs and reaches a lower steady-state value, indi-

	FD-MCNN-BiLSTM [this paper]	LDRSN [13]	TMCCNN [14]	CNN-LSTM [15]
Classifier	Softmax	Softmax	Softmax	Softmax
Loss Function	Cross-entropy	Cross-entropy	Cross-entropy	Cross-entropy
Optimizer	Adam	Adam	Adam	Adam
Learning Rate	0.0001	0.0001	0.0001	0.0001
L2 regularization	10^{-3}	10^{-3}	10^{-3}	10^{-3}
IQ size	2 × 1024	2 × 1024	2 × 1024	2 × 1024
Spectrogram size	3 × 384 × 256	3 × 384 × 256	3 × 384 × 256	3 × 384 × 256
Batch size	16	16	16	16

Tab. 7. Neural network parameter settings.

Project	Content
Operating System	Windows11
CPU	i9-12900
GPU	5060TI
Memory	64G
Deep learning framework	Pytorch 1.12.0
Programming language	Python 3.9

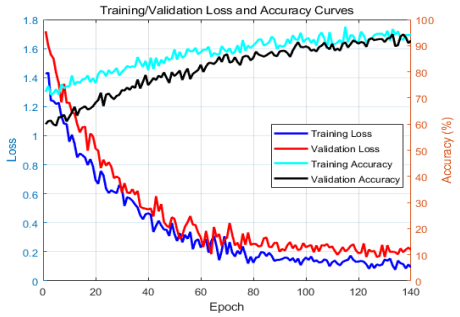
Tab. 8. Description of the experiment environment.

cating higher optimization efficiency and better convergence behavior. In contrast, LDRSN and TMCCNN converge more slowly and show larger fluctuations, suggesting that they are more sensitive to overlapping interference and noise. The training and validation losses of CNN-LSTM both gradually decrease and eventually stabilize, indicating that the model can be optimized relatively smoothly. However, its validation loss remains consistently higher than its training loss, implying limited generalization capability. This limitation may be attributed to its relatively simple front-end feature extraction structure, which is insufficient to capture the multi-scale features and fine-grained differences in mixed signals.

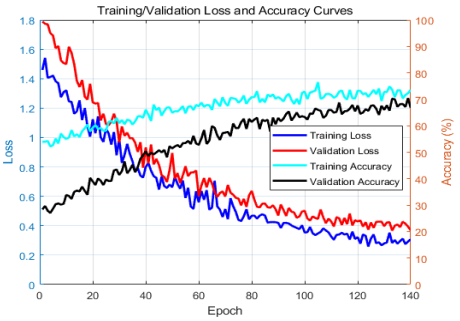
Notably, the loss trajectory of FD-MCNN-BiLSTM remains smooth in the later stages of training, indicating stronger training stability and a lower risk of convergence to suboptimal solutions.

In addition, from the perspective of the accuracy curve, FD-MCNN-BiLSTM shows a faster and more stable increase in training accuracy, rising from about 72% to 94%, with validation accuracy increasing from about 63% to 92%, closely matching the training accuracy. This suggests strong generalization and better convergence compared to other models.

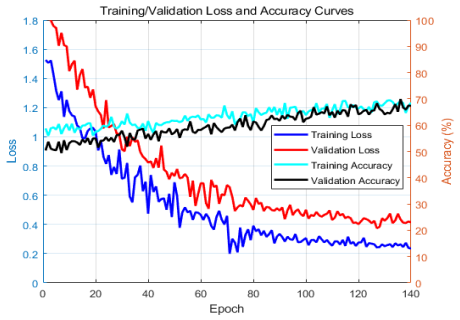
In contrast, training accuracy of LDRSN rises from about 58% to 72%, and validation accuracy from 30% to 69%, with a larger gap between them, indicating poor generalization. TMCCNN shows a gradual increase, with training accuracy rising from about 60% to 70%, and validation accuracy from about 56% to 68%. The flat accuracy curves suggest limited performance in capturing complex features. CNN-LSTM shows a similar trend, with training accuracy rising from about 59% to 78% and validation accuracy from about 48% to 78%, indicating good generalization, though still lagging behind FD-MCNN-BiLSTM in accuracy and efficiency.



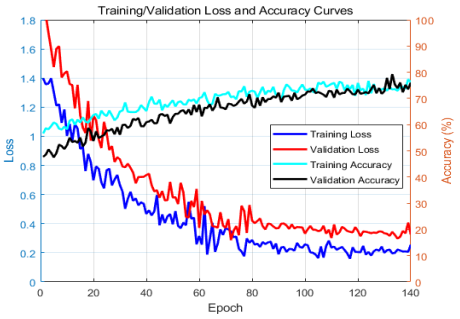
a) FD-MCNN-BiLSTM [this paper]



b) LDRSN [13]



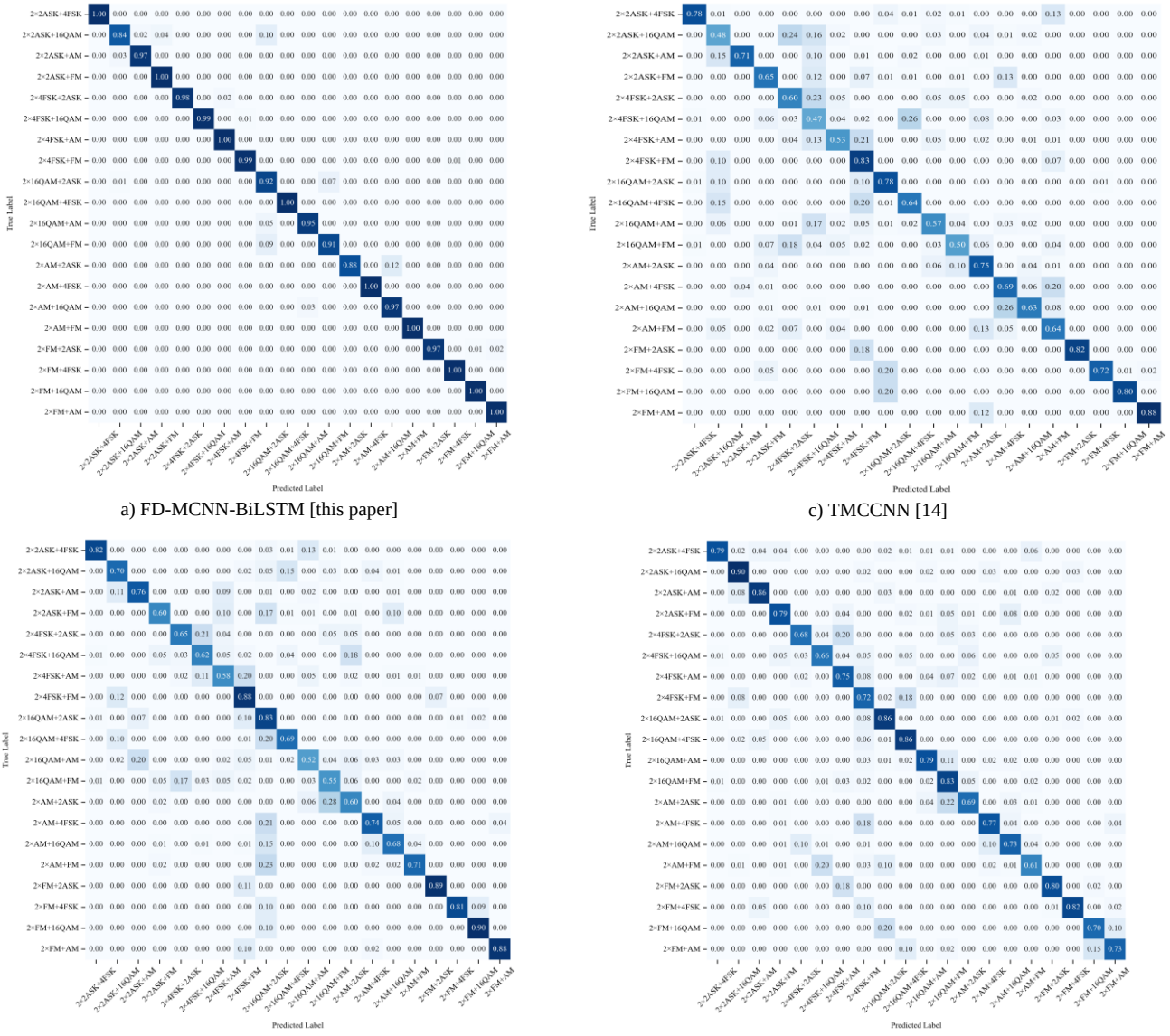
c) TMCCNN [14]



d) CNN-LSTM [15]

Fig. 10. Loss and accuracy curves of four models (30 dB).

Figure 11 shows the confusion matrices of FD-MCNN-BiLSTM, LDRSN, TMCCNN, and CNN-LSTM at 30 dB. As shown in Fig. 11, under the 30 dB condition, the four models exhibit significant differences in their ability to recognize mixed-signal classes. Overall, the proposed FD-MCNN-BiLSTM model demonstrates the most stable performance. Most mixed-signal classes achieve a recognition accuracy of 100%, while the classes 2×4FSK+16QAM and 2×4FSK+FM achieve 99%. These results indicate that the proposed framework effectively exploits multi-domain feature disentanglement and temporal modeling, yielding strong discriminative capability for typical mixed-signal



a) FD-MCNN-BiLSTM [this paper]

c) TMCCNN [14]

b) LDRSN [13]

d) CNN-LSTM [15]

Fig. 11. Confusion matrices of four models at 30 dB.

structures. In contrast, the comparison models show considerable misclassification for certain classes. For example, LDRSN achieves only 52% accuracy on 2×16QAM+AM, with 20% of the samples being misclassified as 2×2ASK+AM. TMCCNN attains only 48% accuracy on 2×2ASK+16QAM, with 24% of the samples being confused with 2×4FSK+2ASK. CNN-LSTM achieves only 61% accuracy on 2×AM+FM, with 20% of the samples being misclassified as 2×4FSK+16QAM.

A closer examination of the misclassified samples of FD-MCNN-BiLSTM also reveals several challenging cases. When recognizing the 2×2ASK+16QAM class, the accuracy is 84%, and this class is more likely to be confused with 2×16QAM+2ASK. For the 2×AM+2ASK class, the accuracy is 88%, with misclassifications mainly occurring as 2×AM+16QAM. These errors typically arise when the local time-frequency structures of mixed samples are highly similar and the representation of one component is influenced by the other.

The comparative results of FD-MCNN-BiLSTM and the other three advanced models under different SNR conditions are presented in Tab. 9 and Fig. 12.

Overall, the results consistently show that FD-MCNN-BiLSTM achieves a significant and stable performance advantage.

It is observed in Fig. 12 that the accuracy curves under different SNR conditions show that the recognition performance of all methods generally improves as the SNR increases from 0 dB to 30 dB. Among them, FD-MCNN-BiLSTM consistently achieves the highest accuracy across the entire SNR range and exhibits a more stable upward trend, indicating better robustness to noise and stronger adaptability to different signal conditions.

In particular, the superiority of the proposed method is more evident at SNR levels of 0, 20, and 30 dB. At 0 dB, FD-MCNN-BiLSTM reaches 80.48%, significantly outper-

	FD-MCNN-BiLSTM [this paper]	LDRSN [13]	TMCCNN [14]	CNN-LSTM [15]
Acc (SNR = 0 dB)	**80.48%**	54.97%	58.02%	48.47%
Acc (SNR = 5 dB)	**79.26%**	57.40%	58.39%	48.35%
Acc (SNR = 10 dB)	**80.46%**	56.23%	59.26%	50.20%
Acc (SNR = 15 dB)	**84.35%**	61.67%	62.07%	63.58%
Acc (SNR = 20 dB)	**89.11%**	67.06%	66.32%	71.30%
Acc (SNR = 25 dB)	**92.56%**	68.72%	68.89%	78.22%
Acc (SNR = 30 dB)	**96.85%**	70.37%	69.91%	78.40%

Tab. 9. Accuracy of four models under different SNRs.

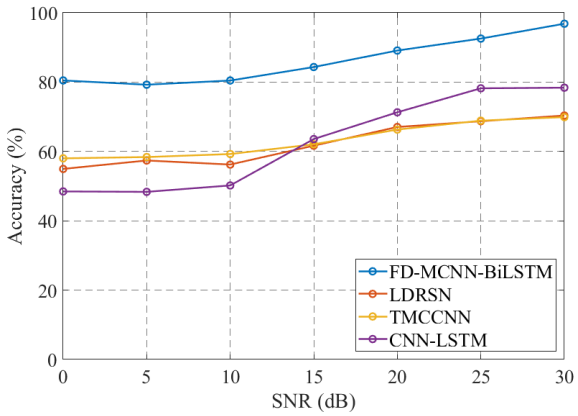


Fig. 12. Accuracy curves of different models under different SNRs.

forming LDRSN (54.97%), TMCCNN (58.02%), and CNN-LSTM (48.47%). At 20 dB, its accuracy increases to 89.11%, remaining clearly higher than that of the comparison methods. When the SNR further increases to 30 dB, the proposed method achieves 96.85%, whereas LDRSN, TMCCNN, and CNN-LSTM reach only 70.37%, 69.91%, and 78.40%, respectively.

These results demonstrate that FD-MCNN-BiLSTM maintains superior recognition performance under both low- and high-SNR conditions, thereby validating the effectiveness of the proposed hierarchical recognition framework and multi-domain feature disentanglement strategy.

4.4 Ablation Studies

To verify the effectiveness of the proposed modules, ablation studies are conducted under the same training protocol and evaluation settings as those used in the main experiments. Unless otherwise specified, each ablation variant retains all remaining components unchanged to ensure a fair comparison. Recognition accuracy and F1-score are reported at SNR levels of 0, 5, 10, 15, 20, 25, and 30 dB, and the corresponding results are summarized in Tabs. 10–15.

(1) Ablation Study on FD-MCNN

This experiment evaluates the contribution of each feature branch in FD-MCNN, including the time-domain branch (T), the frequency-domain branch (F) based on STFT, the modulation-characteristics branch (M) with complex-valued convolutions, and the energy-sensing branch (E).

We compare the full model (T+F+M+E) with variants that remove one branch at a time, while keeping the fusion strategy and the rest of the architecture identical.

As shown in Tab. 10 and Fig. 13, the full four-branch FD-MCNN achieves the best performance across all evaluated SNRs, indicating that multi-domain feature disentanglement provides complementary cues for mixed-signal recognition. Removing any single branch results in a noticeable accuracy drop, confirming that each pathway contributes meaningfully to the final decision.

More specifically, removing the time-domain branch causes the largest performance degradation, especially at 30 dB, suggesting that temporal correlation patterns are critical under severe signal mixing, where weak components are more easily masked. In contrast, removing the

T	✓	×	✓	✓	✓
F	✓	✓	×	✓	✓
M	✓	✓	✓	×	✓
E	✓	✓	✓	✓	×
Acc (SNR = 0 dB)	**80.48%**	58.30%	70.82%	65.57%	71.22%
Acc (SNR = 5 dB)	**79.26%**	60.91%	73.65%	66.67%	72.50%
Acc (SNR = 10 dB)	**80.46%**	62.17%	70.21%	69.00%	76.27%
Acc (SNR = 15 dB)	**84.35%**	61.88%	71.02%	68.91%	78.90%
Acc (SNR = 20 dB)	**89.11%**	64.05%	74.36%	64.38%	77.11%
Acc (SNR = 25 dB)	**92.56%**	64.27%	73.46%	67.09%	78.26%
Acc (SNR = 30 dB)	**96.85%**	69.14%	79.56%	72.39%	84.28%

Tab. 10. Ablation study results of the FD-MCNN branches.

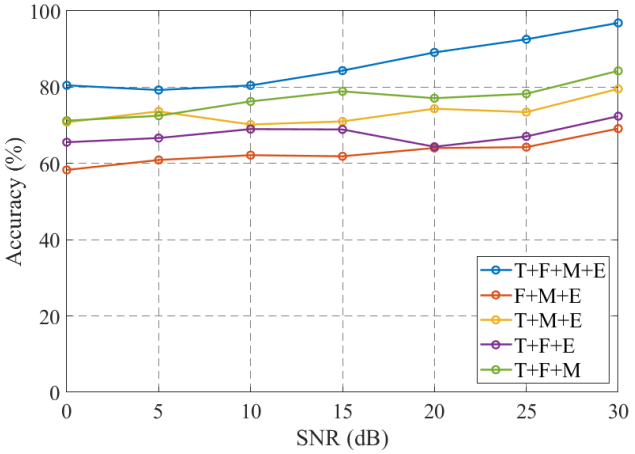


Fig. 13. Accuracy curve under different SNRs.

T	✓	×	✓	✓	✓
F	✓	✓	×	✓	✓
M	✓	✓	✓	×	✓
E	✓	✓	✓	✓	×
F1-score (SNR = 0 dB)	**80.20%**	54.48%	67.21%	61.75%	71.03%
F1-score (SNR = 5 dB)	**79.29%**	55.40%	69.97%	62.77%	72.47%
F1-score (SNR = 10 dB)	**80.44%**	57.95%	69.08%	64.03%	76.21%
F1-score (SNR = 15 dB)	**84.37%**	60.27%	68.78%	65.11%	78.93%
F1-score (SNR = 20 dB)	**89.10%**	61.60%	71.40%	63.26%	77.15%
F1-score (SNR = 25 dB)	**92.57%**	61.67%	71.25%	63.50%	78.33%
F1-score (SNR = 30 dB)	**96.87%**	62.35%	74.67%	69.09%	84.40%

Tab. 11. F1-score results for FD-MCNN branch ablation.

energy-sensing branch leads to a relatively smaller decrease, implying that its contribution can be partially compensated by the remaining branches. Overall, these results support the effectiveness of integrating time-, frequency-, modulation-, and energy-domain information to improve recognition robustness under mixed and interference-rich conditions.

The corresponding F1-score results are shown in Tab. 11. As shown in Tab. 11, the complete four-branch FD-MCNN achieves F1-scores of 80.20%, 79.29%, 80.44%, 84.37%, 89.10%, 92.57%, and 96.87% at 0, 5, 10, 15, 20, 25, and 30 dB, respectively, consistently outperforming all ablation variants. Since the F1-score jointly reflects precision and recall, these results further indicate that the complete model provides not only stronger recognition capability but also better prediction balance across different classes.

A comparison among the ablation variants shows that removing the time-domain branch has the most detrimental effect, with the F1-score ranging only from 54.48% to 62.35% across all SNR conditions, which is substantially lower than that of the other variants. This result suggests that the time-domain branch provides essential cues for capturing transient waveforms, local temporal variations, and I/Q dynamic structures in mixed signals.

By contrast, although the model performance also declines after removing the frequency-domain branch or the energy-sensing branch, it remains noticeably better than that of the variant without the time-domain branch. When the modulation-characteristics branch is removed, the F1-score ranges from 61.75% to 69.09%, showing a clear overall degradation. This demonstrates that the modulation-domain branch plays an important role in preserving I/Q phase relationships and enhancing modulation-pattern representation.

(2) Ablation Study on CSTA

This experiment investigates whether the proposed CSTA mechanism enhances feature disentanglement and improves recognition robustness for overlapping mixed signals. The full model is compared with a variant in which CSTA is removed, while all other components are kept unchanged.

CSTA	✓	×
Acc (SNR = 0 dB)	**80.48%**	64.19%
Acc (SNR = 5 dB)	**79.26%**	64.30%
Acc (SNR = 10 dB)	**80.46%**	66.82%
Acc (SNR = 15 dB)	**84.35%**	74.67%
Acc (SNR = 20 dB)	**89.11%**	80.57%
Acc (SNR = 25 dB)	**92.56%**	84.89%
Acc (SNR = 30 dB)	**96.85%**	86.25%

Tab. 12. Ablation study results of the CSTA.

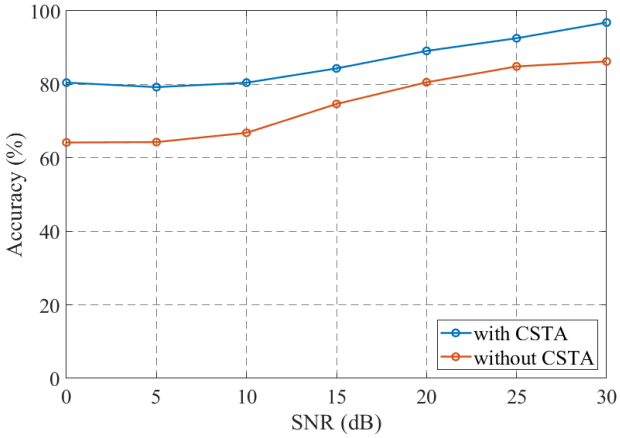


Fig. 14. Accuracy curves of CSTA ablation under different SNRs.

The results are summarized in Tab. 12 and Fig. 14. As shown in Tab. 12 and Fig. 14, incorporating CSTA consistently improves recognition accuracy at all evaluated SNR levels. The gains are more pronounced at 0, 5, and 10 dB, indicating that attention-based reweighting is particularly beneficial under low-SNR conditions, where interference is stronger and weak components are more easily suppressed.

This observation supports the design motivation of CSTA: by adaptively emphasizing informative channel-wise and spatiotemporal features while suppressing irrelevant activations, the model learns more discriminative representations for mixed signals.

The corresponding F1-scores are reported in Tab. 13. Table 13 shows that the model with CSTA consistently outperforms the variant without CSTA under all SNR conditions. As the SNR increases from 0 dB to 30 dB, the F1-score of the model with CSTA rises from 80.20% to 96.87%, whereas that of the model without CSTA increases

CSTA	✓	×
F1-score (SNR = 0 dB)	**80.20%**	61.25%
F1-score (SNR = 5 dB)	**79.29%**	62.44%
F1-score (SNR = 10 dB)	**80.44%**	62.97%
F1-score (SNR = 15 dB)	**84.37%**	68.88%
F1-score (SNR = 20 dB)	**89.10%**	75.30%
F1-score (SNR = 25 dB)	**92.57%**	78.29%
F1-score (SNR = 30 dB)	**96.87%**	81.43%

Tab. 13. Ablation study results of the CSTA.

only from 61.25% to 81.43%. These results indicate that CSTA can stably enhance recognition performance under varying noise conditions. More specifically, the gain introduced by CSTA is particularly evident under low-SNR conditions. For example, at 0 dB, incorporating CSTA improves the F1-score by 18.95 percentage points; even at 30 dB, it still yields an improvement of 15.44 percentage points. This demonstrates that CSTA not only strengthens the extraction of weak yet informative features but also improves model robustness in complex noise environments.

Overall, the results verify the effectiveness of CSTA in feature enhancement and discriminative information modeling.

(3) Ablation Study on BiLSTM

This experiment investigates the necessity of the BiLSTM module for capturing temporal dependencies in mixed-signal recognition. To this end, BiLSTM is replaced with a temporal pooling layer followed by fully connected layers, while FD-MCNN and CSTA are kept unchanged. The results are reported in Tab. 14 and Fig. 15.

As shown in Tab. 14 and Fig. 15, removing the BiLSTM degrades recognition accuracy across all SNR levels, confirming that explicit temporal modeling contributes substantially to performance improvement.

The advantage of BiLSTM is more pronounced at high SNR levels (20, 25, and 30 dB), suggesting that bidirectional contextual modeling helps distinguish mixed modulation patterns by exploiting longer-range temporal dependencies. This capability is particularly important in mixed-signal scenarios, where instantaneous features alone may be insufficient because of interference, noise, and partial masking.

BiLSTM	✓	×
Acc (SNR = 0 dB)	**80.48%**	66.38%
Acc (SNR = 5 dB)	**79.26%**	69.59%
Acc (SNR = 10 dB)	**80.46%**	68.27%
Acc (SNR = 15 dB)	**84.35%**	70.04%
Acc (SNR = 20 dB)	**89.11%**	72.80%
Acc (SNR = 25 dB)	**92.56%**	76.25%
Acc (SNR = 30 dB)	**96.85%**	81.77%

Tab. 14. Ablation study results of the BiLSTM.

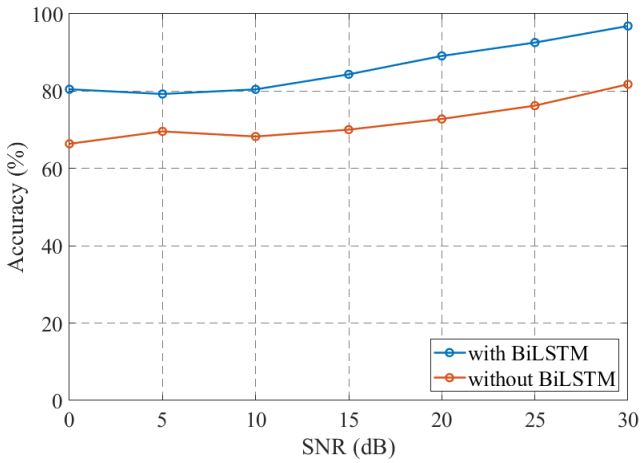


Fig. 15. Accuracy curves of BiLSTM ablation under different SNRs.

BiLSTM	✓	×
F1-score (SNR = 0 dB)	**80.20%**	61.40%
F1-score (SNR = 5 dB)	**79.29%**	62.67%
F1-score (SNR = 10 dB)	**80.44%**	62.96%
F1-score (SNR = 15 dB)	**84.37%**	65.34%
F1-score (SNR = 20 dB)	**89.10%**	68.67%
F1-score (SNR = 25 dB)	**92.57%**	72.34%
F1-score (SNR = 30 dB)	**96.87%**	77.39%

Tab. 15. Ablation study results of the BiLSTM.

The corresponding F1-scores are listed in Tab. 15. Table 15 shows that the model with BiLSTM consistently outperforms the variant without BiLSTM under all SNR conditions. As the SNR increases from 0 dB to 30 dB, the F1-score of the model with BiLSTM improves from 80.20% to 96.87%, whereas that of the model without BiLSTM increases only from 61.40% to 77.39%. This indicates that the BiLSTM module effectively enhances the overall recognition capability of the model.

Across different SNR conditions, the performance gain brought by BiLSTM remains consistently significant. At 0 dB, incorporating BiLSTM improves the F1-score by 18.80 percentage points, while at 30 dB the improvement further reaches 19.48 percentage points. These results demonstrate that BiLSTM can effectively exploit both forward and backward contextual information to model temporal dependencies in mixed signals, thereby improving the discriminative capability of the model for different modulation components.

Overall, the BiLSTM module plays a crucial role in enhancing mixed-signal recognition performance, which verifies its effectiveness in temporal dependency modeling.

(4) Summary of Findings

Overall, the ablation results consistently demonstrate that all three components contribute positively to the final performance.

The four-branch FD-MCNN provides complementary multi-domain information, CSTA enhances discriminative feature representation and robustness under noise, and BiLSTM further improves recognition by modeling temporal dependencies. These findings jointly validate the design rationale of FD-MCNN-BiLSTM and help explain its superior performance over TMCCNN, LDRSN, and CNN-LSTM in the main experiments.

5. Conclusion

This paper presents FD-MCNN-BiLSTM, a hierarchical recognition framework for mixed signals. By systematically extracting and fusing time-domain, frequency-domain, modulation characteristics, and energy cues, the

proposed method enables layered identification of dominant and weak components in realistic mixtures. The key novelty lies in the FD-MCNN coupled with a BiLSTM for temporal dependency modeling, while the CSTA mechanism further strengthens feature disentanglement and weak-signal sensitivity under masking effects.

Experimental results on a self-constructed mixed-signal dataset verify the effectiveness of the proposed model. Compared with TMCCNN, LDRSN and CNN-LSTM, FD-MCNN-BiLSTM achieves higher recognition accuracy at different SNR conditions (0, 5, 10, 15, 20, 25, 30 dB), demonstrating consistent improvements across medium-to-high SNR regimes and enhanced robustness under mixed and overlapping conditions.

Overall, these findings indicate that multi-domain feature disentanglement combined with hierarchical strong-to-weak recognition is a practical and reliable strategy for modulation recognition, with meaningful implications for spectrum sensing and intelligent receivers in interference-rich electromagnetic environments.

Despite the encouraging performance, several limitations remain. First, the current evaluation is conducted on the constructed dataset and the considered mixture settings; generalization to broader channel conditions (e.g., fading diversity, carrier frequency offsets, sampling-rate mismatch, and receiver nonidealities) requires further verification. Second, the multi-column architecture and attention modules introduce additional computational overhead, which may hinder real-time deployment on resource-constrained platforms. Third, the weak-signal enhancement strategy could be further improved for extremely low-SNR cases or mixtures involving more than two concurrent components.

Future work will focus on (i) extending the framework to multi-source mixtures and more diverse waveform families, (ii) improving robustness via domain adaptation and self-/semi-supervised learning with limited labels, (iii) explicitly modeling practical impairments for field deployment, and (iv) reducing complexity through lightweight design and pruning/quantization for real-time implementation.

Acknowledgments

The work was supported in part by the National Natural Science Foundation of China, Project under Grant 62271419 and Grant 62361136810; in part by the CETC Key Laboratory of ElectroMagnetic Operation and Application, under Grant No. JTKLAB-2024008.

References

[1] YU, H., YAN, X., LIU, S., et al. Radar emitter multi-label recognition based on residual network. *Defence Technology*, 2022, vol. 18, no. 3, p. 410–417. DOI: 10.1016/J.DT.2021.02.005

[2] ELSAYED, M., EL-BANNA, A. A. A., DOBRE, O. A., et al. Machine learning-based self-interference cancellation for full-duplex radio: Approaches, open challenges, and future research directions. *IEEE Open Journal of Vehicular Technology*, 2024, vol. 5, p. 21–47. DOI: 10.1109/OJVT.2023.3331185

[3] SATHAYE, H., MOTALLEBIGHOMI, M., RANGANATHAN, A. Galileo-SDR-SIM: An open-source tool for generating Galileo satellite signals. In *Proceedings of the 36th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2023)*. Denver (Colorado), 2023, p. 3470–3480. DOI: 10.33012/2023.19254

[4] CAO, J., YANG, Z., TIAN, S., et al. Biprobes blade tip timing method for frequency identification based on active aliasing time-delay estimation and dealiasing. *IEEE Transactions on Industrial Electronics*, 2023, vol. 70, no. 2, p. 1939–1948. DOI: 10.1109/TIE.2022.3165252

[5] DENG, J., SUN, Z., LI, X., et al. ASC-BSS-based parameter estimation method for multiple LFM pulses with aliasing effect from passive radar. *IEEE Transactions on Aerospace and Electronic Systems*, 2025, vol. 61, no. 2, p. 5132–5144. DOI: 10.1109/TAES.2024.3516706

[6] HAJEK, K., KOHL, Z. Multiple aliasing of windowed real-valued signal as a cause of accuracy limitation of DFT methods. *IEEE Access*, 2025, vol. 13, p. 6515–6526. DOI: 10.1109/ACCESS.2025.3526325

[7] 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

[8] POLAK, L., TURAK, S., SOTNER, R., et al. Exploring deep learning architectures for RF signal classification. In *35th International Conference Radioelektronika (RADIOELEKTRONIKA)*. Czech Republic, 2025, p. 1–6. DOI: 10.1109/RADIOELEKTRONIKA65656.2025.11008396

[9] PAN, Z., WANG, B., ZHANG, R., et al. MIML-GAN: A GAN-based algorithm for multi-instance multi-label learning on overlapping signal waveform recognition. *IEEE Transactions on Signal Processing*, 2023, vol. 71, p. 859–872. DOI: 10.1109/TSP.2023.3242091

[10] YANG, W., REN, K., DU, Y., et al. Modulation recognition method of mixed signals based on cyclic spectrum projection. *Scientific Reports*, 2023, vol. 13, p. 1–18. DOI: 10.1038/s41598-023-48467-w

[11] LI, Y. Single-channel time-frequency overlapped signal separation algorithm research (in Chinese). *Ph.D. Dissertation*. University of Electronic Science and Technology of China, 2021.

[12] YANG, Y. Key technologies for blind separation and modulation recognition of single-channel overlapped signals (in Chinese). *Ph.D. Dissertation*. Beijing University of Posts and Telecommunications (China), 2022.

[13] LIU, J., WEI, X., FAN, J., et al. Mixed signal modulation recognition method based on temporal depth residual shrinkage network (in Chinese). *Telecommunications Science*, 2024, vol. 40, no. 10, p. 27–38. DOI: 10.11959/j.issn.1000-0801.2024207

[14] KANG, Y., CHEN, K., CHENG, C., et al. Mixed signal modulation classification based on deep convolutional neural networks. In *2024 3rd International Conference on Electronics and Information Technology (EIT)*. Chengdu (China), 2024, p. 655–659. DOI: 10.1109/EIT63098.2024.10762196

[15] XU, J., LIN, Z. Modulation and classification of mixed signals based on deep learning. *arXiv preprint*, 2022, p. 1–20. DOI: 10.48550/arXiv.2205.09916

[16] LVLLN. *GitHub: Hierarchical-recognition-model*. March 2026. [Online]. Available: https://github.com/Yoihhh/Hierarchical-recognition-model

[17] LIU, X., LI, J., JIN, C. T., et al. Wireless signal representation techniques for automatic modulation classification. *IEEE Access*, 2022, vol. 10, p. 84166–84187. DOI: 10.1109/ACCESS.2022.3197224

[18] AKINSANMI, O., GBANGBALA, U., AYEORIBE, O. P. Analog and digital modulation techniques in communication electronics - noise performance comparison. *International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering*, 2025, vol. 4, no. 1, p. 1–10. DOI: 10.5281/zenodo.17591387

[19] NI, C., LIN, T., XING, J., et al. A time-frequency analysis of non-stationary signals using variation mode decomposition and synchrosqueezing techniques. In *Proceedings of the 2019 Prognostics and System Health Management Conference (PHM-Qingdao)*. Qingdao (China), 2019, p. 1–6. DOI: 10.1109/PHM-Qingdao46334.2019.8943036

[20] GUY, Y., WANGY, Y., ADEBISIZ, B., et al. Blind signal recognition method of STBC based on multi-channel convolutional neural network. In *2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall)*. London (UK), 2022, p. 1–5. DOI: 10.1109/VTC2022-Fall57202.2022.10012817

[21] WANG, Z., FU, H., HU, C., et al. Multi-source partial discharge pattern recognition in GIS based on Grabcut-MCNN. *Journal of Measurements in Engineering*, 2024, vol. 13, no. 1, p. 89–104. DOI: 10.21595/JME.2024.24274

[22] SINSOMBOONTHONG, S. Performance comparison of new adjusted min-max with decimal scaling and statistical column normalization methods for artificial neural network classification. *International Journal of Mathematics and Mathematical Sciences*, 2022, no. 1, p. 1–9. DOI: 10.1155/2022/3584406

[23] SRIVASTAVA, H. M., SHAH, F. A., QADRI, H. L., et al. Quadratic-phase Hilbert transform and the associated Bedrosian theorem. *Axioms*, 2023, vol. 12, no. 2, p. 1–15. DOI: 10.3390/axioms12020218

[24] KUNDU, M. Different types of Plancherel's theorems for square integrable functions associated with quaternion offset linear canonical transforms. *Journal of the Franklin Institute*, 2025, vol. 362, no. 7, p. 1–21. DOI: 10.1016/j.jfranklin.2025.107649

[25] HASSANZADEH, M., SHAHRRAVA, B. Linear version of Parseval's theorem. *IEEE Access*, 2022, vol. 10, p. 27230–27241. DOI: 10.1109/ACCESS.2022.3157736

[26] RATHOD, D., SONDAGAR, H., SHAH, V., et al. Optimizing TimeSformer for video analysis and action recognition. In *International Conference on Artificial Intelligence and Machine Vision (AIMV)*, Gandhinagar (India), 2025, p. 1–6. DOI: 10.1109/AIMV66517.2025.11203298

[27] WANG, Y., WANG, J., BAI, X. Three-stream fusion networks for student engagement recognition based on TimeSformer. In *International Conference on Artificial Intelligence and Computer Engineering (ICAICE 2022) - Proc. of SPIE*. Wuhan (China), 2023, vol. 12610, p. 1–7. DOI: 10.1117/12.2671122

[28] HE, Y., CAO, N., HU, C., et al. A CNN-BiLSTM-based deep learning model for CPM signal detection. *Electronics Letters*, 2025, vol. 61, no. 1, p. 1–6. DOI: 10.1049/ELL2.70502

[29] IBRAHIM, M., BADRAN, K. M., ESMAT, H. A. Artificial intelligence-based approach for univariate time-series anomaly detection using hybrid CNN-BiLSTM model. In *13th International Conference on Electrical Engineering (ICEENG)*. Cairo (Egypt), 2022, p. 129–133. DOI: 10.1109/ICEENG49683.2022.9781894

[30] NEILI, Z., SUNDARAJ, K. Addressing varying lengths in PCG signal classification with BiLSTM model and MFCC features. In *8th International Conference on Image and Signal Processing and their Applications (ISPA)*. Biskra (Algeria), 2024, p. 1–5. DOI: 10.1109/ISPA59904.2024.10536851

[31] BUCKLAND, M., GEY, F. The relationship between recall and precision. *Journal of the American Society for Information Science*, 1994, vol. 45, no. 1, p. 12–19. DOI: 10.1002/(SICI)1097-4571(199401)45:1<12::AID-ASI2>3.0.CO;2-L

[32] LIN, T. Y., GOYAL, P., GIRSHICK, R., et al. Focal loss for dense object detection. In *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*. Venice (Italy), 2017, p. 2999–3007. DOI: 10.1109/ICCV.2017.324