A Signal Cognition Method Based on Multi-View Evidence and Hierarchical Probabilistic Outputs Without Prior Information

Xiaolin Xiu

Independent Researcher
ORCID: https://orcid.org/0009-0001-0946-7067

Abstract

Signal cognition without prior information is often required in electromagnetic situational awareness, cognitive radio, and signals intelligence (SIGINT). This paper proposes a blind signal cognition method based on multi-view evidence and hierarchical probabilistic outputs. A fixed-length in-phase/quadrature (I/Q) data segment is used as one cognition window. Multidimensional physical features are extracted from the time, frequency, and time-frequency domains and at different time scales. A staged multi-layer perceptron (MLP) then produces support probabilities for the reference background, signal process, basic signal form, signal family, and fine-grained signal type. Complete upstream probability vectors are also passed to later stages, allowing the model to retain broader cognition results when the evidence is insufficient for a fine-grained distinction. Simulations involving 73 signal types show that, at signal-to-noise ratios (SNRs) of 0 dB or higher, the average recognition rates for basic signal form, signal family, and fine-grained type reach 92.40%, 87.63%, and 87.15%, respectively. The results indicate that combining multi-view evidence with a hierarchical soft cascade provides a feasible approach to multi-class signal cognition when no prior information about the input signal is available.

Keywords: blind signal cognition; multi-view features; hierarchical probabilistic model; soft cascade

1 Introduction

Signal cognition is widely used in electromagnetic situational awareness, non-cooperative signal analysis, and electronic warfare. In complex environments, it can also provide preliminary information for interference mitigation and the selection of subsequent receiver chains in communication, navigation, and radar systems. Two basic tasks are to detect a signal against background noise and then determine its form and structure. Unlike communication, navigation, and radar receivers designed for known signal systems, a cognition system is not given the signal class, SNR, occupied bandwidth, or modulation parameters of the current input at inference time. It is therefore difficult to construct a signal-specific matched filter in advance to obtain processing gain.

Early research on automatic signal recognition focused mainly on modulation classification. Dobre et al. divided classical methods into likelihood-based and feature-based approaches. Likelihood-based methods calculate the observation likelihood under candidate signal models and generally require assumptions about, or estimates of, unknown parameters such as noise, carrier, phase, and timing. Feature-based methods extract statistics such as higher-order cumulants, cyclostationary features, amplitude, phase, and instantaneous frequency from received data, and then apply decision rules or conventional machine-learning classifiers [1]. For example, Hussain et al. constructed modulation-discriminative features using optimized linear combinations of higher-order signal cumulants [2]. Although these methods have relatively clear physical interpretations, their performance often depends on the signal model, feature selection, and parameter-estimation accuracy. Their performance may be limited at low SNRs, when the number of cognition classes increases, or when channel conditions change.

In recent years, deep learning has shifted signal cognition from decisions based on handcrafted features toward data-driven representation learning. O'Shea and Hoydis demonstrated that a convolutional neural network (CNN) can learn discriminative modulation features directly from raw I/Q samples [3]. Qu et al. combined time-frequency analysis with a CNN to represent structural differences among radar intrapulse modulations using time-frequency images [4]. Zhang et al. further fused I/Q, amplitude-phase, and multiple spectral representations, showing that different signal features can provide complementary information [5]. These studies established several common types of input, including raw samples, handcrafted statistical features, spectra, and time-frequency images, and substantially improved signal-recognition performance under complex background conditions. Nevertheless, such evidence fusion is still generally used for a predefined closed-set classification task that produces one final label from a fixed-length segment containing a detected or extracted single signal.

Other studies have extended conventional narrowband single-signal classification to more complex spectrum-cognition problems. WBSig53, proposed by Boegner et al., extends the task to multi-signal detection, time-frequency localization, and modulation-family recognition in wideband observations [6]. Shi et al. discussed changes in signal types, unknown signals, spoofing signals, and multi-signal superposition in dynamic spectrum environments [7]. To handle uncertainty, open-set automatic modulation recognition attempts to avoid forced closed-set decisions by detecting unknown modulation classes [8]. Early-exit networks determine whether inference should terminate according to sample difficulty, trading recognition accuracy against computational cost [9]. In more general machine-learning research, hierarchical selective classification allows a model to return a less specific prediction when the evidence is insufficient, thereby avoiding risky fine-grained decisions [10].

These lines of work have advanced wideband detection, multi-domain fusion, unknown-class recognition, computational early exit, and hierarchical output, but they address related rather than equivalent problems. Wideband detection primarily asks when and at what frequency a signal is present. Multi-domain fusion generally remains aimed at improving final classification accuracy. Open-set methods focus on whether an input belongs to a known class. Early-exit methods mainly decide whether further computation is required. Hierarchical selective classification has not yet been organized into a unified processing framework for the physical meanings of communication, navigation, and radar signals. Moreover, although many studies do not provide the cognition model with modulation labels or some channel parameters, they often assume that the target signal has already been detected, extracted, and normalized before the model input is formed. These studies also commonly adopt a closed-set setting in which the test classes largely agree with the taxonomy predefined during training. This motivates the study of how multi-domain evidence, the absence of prior input information, and hierarchical cognition can be combined in one processing chain for continuous I/Q streams.

This paper presents an exploratory study of general-purpose signal cognition. The proposed multi-view signal cognition processor treats a continuous I/Q stream as its engineering input and a finite-length

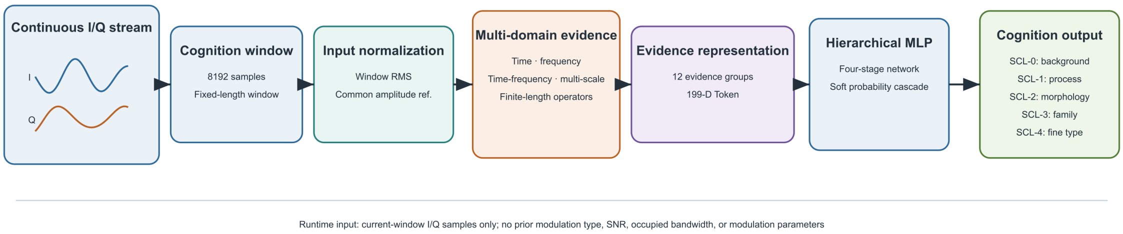

Figure 1: Overall data flow of the multi-view signal cognition processor

segment as one cognition window. It computes physically meaningful and complementary evidence in the time, frequency, and time-frequency domains. A hierarchical neural network then produces cognition results in stages: background and signal-process screening, basic signal form, signal family, and fine-grained signal type. The aim is not to optimize one classification metric for a specific application, but to examine the feasibility and limits of a general processing chain that combines multi-domain observation with hierarchical cognition and avoids unsupported fine-grained decisions. The study also assesses the engineering feasibility of implementing the processor on a field-programmable gate array (FPGA) or digital signal processor (DSP) in terms of feature-extraction cost, model size, fixed-point quantization, and streaming-processing results.

2 Signal Cognition Based on Multi-View Evidence and a Hierarchical Neural Network

The signals considered in this study include common communication and navigation signals, radar intra-pulse modulation signals, and representative forms such as multicarrier, spread-spectrum, pulsed, burst, and swept-frequency signals. Without prior knowledge of the input signal type, SNR, occupied bandwidth, or modulation parameters, the processor accepts a complex I/Q stream and sequentially performs input normalization, multi-domain evidence extraction, cognition feature-vector (Token) construction, MLP inference, and probabilistic output within a fixed-length window. The processor can therefore be divided into three components: multi-view feature processing, a hierarchical cognition model, and cascaded outputs. The overall data flow is shown in Fig. 1.

2.1 Multi-View Feature Extraction and Processing

Let the complex baseband observation within one cognition window be $x[n] = I[n] + \mathrm{j}Q[n]$, $n = 0, 1, \ldots, N-1$. To balance the ability to observe slowly varying signal structures against front-end computational cost, this study uses $N = 8192$ and sets the ratio of occupied input-signal bandwidth to sampling rate, B/F_s, to 0.1-0.45.

The entire input segment is first normalized by its root mean square:

$$\widetilde{x}[n] = \frac{x[n]}{\sqrt{\frac{1}{N}\sum_{r=0}^{N-1}|x[r]|^2 + \varepsilon}},$$

where ε is a small constant that prevents numerical anomalies for zero-energy inputs. Based on $\widetilde{x}[n]$, a set of clearly defined finite-length operations extracts evidence for physical properties such as envelope fluctuations, phase continuity, spectral-line structure, occupied bandwidth, time-frequency trajectories, pulse periodicity, and delay correlation. Time-domain evidence mainly describes the envelope, phase variation, energy distribution, and delay correlation. Let $a[n] = |\widetilde{x}[n]|$. Its coefficient of variation is

$$C_a = \frac{\sigma_a}{\mu_a + \varepsilon},$$

where μ_a and σ_a are the mean and standard deviation, respectively, of the envelope sequence $a[n]$ within the cognition window. This quantity characterizes envelope fluctuations relative to the mean amplitude.

The phase increment and instantaneous frequency between adjacent samples are defined as

$$\Delta\phi[n] = \arg(\widetilde{x}[n]\widetilde{x}^*[n-1]), \qquad f_{\mathrm{i}}[n] = \frac{F_s}{2\pi}\Delta\phi[n],$$

where $n = 1, 2, \ldots, N-1$ and F_s is the sampling rate. The distribution width, continuity, stepwise behavior, and frequency components of $f_{\mathrm{i}}[n]$ provide evidence for distinguishing continuous phase variation, discrete frequency states, and frequency-swept processes.

Similarly, the normalized delay-correlation measure is

$$\rho_\ell = \frac{\left|\sum_{n=\ell}^{N-1} \widetilde{x}[n]\widetilde{x}^*[n-\ell]\right|}{\sum_{n=0}^{N-1} |\widetilde{x}[n]|^2 + \varepsilon},$$

where ℓ is the delay in samples, the superscript * denotes complex conjugation, and the summation spans the valid overlap between the original and delayed sequences. This quantity represents short-term coherence and periodic structure.

Frequency-domain and time-frequency evidence is obtained using windowed Fourier transforms. For the mth short-time frame,

$$X_m[k] = \sum_{n=0}^{L-1} \widetilde{x}[n+mH]w[n] \, \exp\left(-\mathrm{j}\frac{2\pi kn}{L}\right), \qquad P_m[k] = |X_m[k]|^2,$$

where m is the short-time frame index, $k = 0, 1, \ldots, L-1$ is the frequency-bin index, L is the frame length, H is the frame hop, and $w[n]$ is a Hann window. The base trajectory analysis uses $L = 128$ and $H = 32$. A single-frame or aggregated multi-frame power spectrum can be used to calculate spectral-peak prominence, spectral flatness, band occupancy, comb-line spacing, multicarrier persistence, and related evidence. For example, let $P[k]$, $k = 0, 1, \ldots, K-1$, denote the single-frame or aggregated multi-frame power spectrum under consideration and let $p[k] = P[k]/\sum_{r=0}^{K-1} P[r]$. The normalized spectral entropy is

$$H_{\mathrm{s}} = -\frac{\sum_k p[k] \log p[k]}{\log K},$$

where K is the number of valid frequency bins. Spectral entropy increases as the spectrum approaches a uniform distribution.

To describe the temporal evolution of signal structure, the spectral centroid and spectral width can be calculated from the short-time power spectrum:

$$c_m = \frac{\sum_k k P_m[k]}{\sum_k P_m[k]}, \qquad b_m = \frac{1}{L/2}\sqrt{\frac{\sum_k (k-c_m)^2 P_m[k]}{\sum_k P_m[k]}}.$$

Here, c_m and b_m are the spectral centroid of the mth frame and the spectral width normalized by half the frequency-band width, respectively; the sums span the frequency bins used in the calculation. The sequences $\{c_m, b_m\}$ yield trajectory features such as centroid slope, linearity, number of jumps, dwell ratio, and bandwidth stability.

The processor front end observes local variations at different scales using 8-frame and 32-frame views. It also divides the 8192-sample window into 32 subwindows of 256 samples and accumulates cross-window evidence for spectral shape, multicarrier structure, pulse activity, and swept-frequency trajectories. All these calculations are performed within the same cognition window. After evaluation, 199 feature fields in 12 groups are retained, as listed in Table 1.

Table 1. Multi-view evidence groups, dimensions, and principal observation targets

Evidence group	Dimensions	Principal observation targets
Static spectrum and envelope statistics	32	Envelope fluctuation, kurtosis, short-time energy, spectral entropy, band occupancy, and short-delay correlation
8-frame trajectories	12	Spectral peaks, centroids, bandwidths, jumps, and dwell over a shorter time scale
32-frame trajectories	12	Spectral-peak motion, linearity, and stability over a longer time scale
Semantic frame summaries	18	Framewise energy, bandwidth, phase increment, correlation, and ridge statistics
Spectral-line and multicarrier evidence	17	Narrowband spectral lines, comb-line spacing, spectral-peak persistence, and multicarrier consistency
Instantaneous-frequency statistics	12	Instantaneous-frequency distribution, continuity, stepwise behavior, and rate of change
Angular modulation and global instantaneous-frequency evidence	9	Envelope-phase coupling, sideband asymmetry, and frequency smoothness
Angular time-frequency structure	13	Time-frequency ridge width, centroid motion, jumps, curvature, and continuity
Multicarrier structure	10	Active frequency bins, spectral-concentration quality, flatness, and mask persistence
Pulse, burst, and sweep structure	18	Duty cycle, silent intervals, edges, periodicity, intrapulse coherence, and ridge slope
Spread-spectrum, pseudo-random noise (PRN), and delay evidence	7	Correlation peak and sidelobes, wideband occupancy, binary offset carrier (BOC) splitting, and phase-run structure
32-subwindow accumulation and multi-aperture evidence	39	Cross-window envelope, correlation, spectral shape, multicarrier structure, activity state, and time-frequency trajectories

The 199-dimensional Token forms the processor's complete feature representation. When the Token is further compressed to 160 dimensions, simulation shows a maximum decrease of 0.43 percentage points in average recognition rate and a decrease of 5.59 percentage points in the correct support probability of the weakest class.

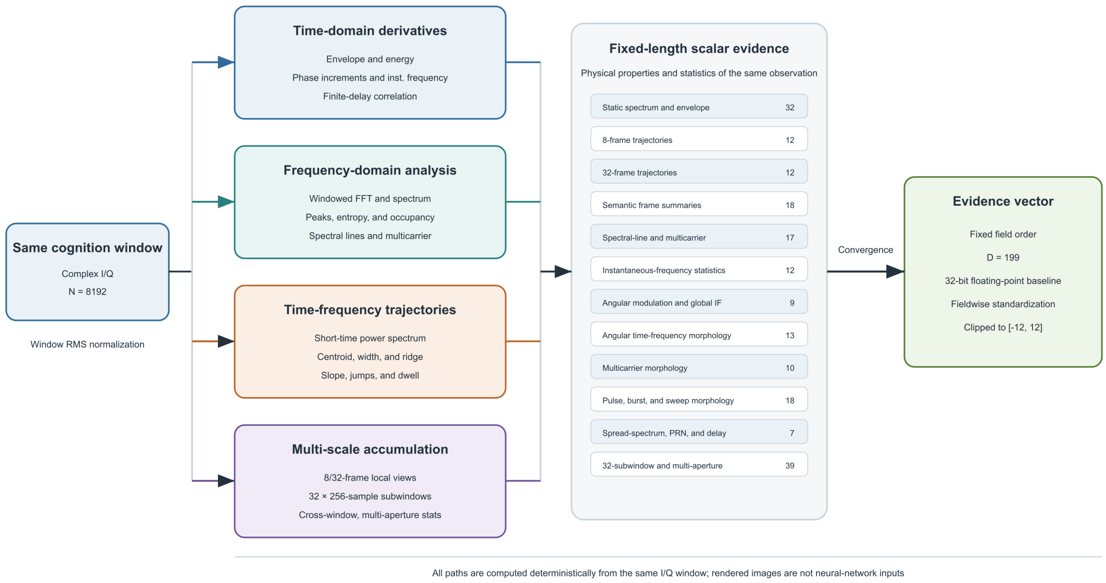

Figure 2: Multi-domain feature extraction and evidence convergence

The four evidence paths and their convergence are shown in Fig. 2.

The scalar evidence obtained through multi-domain feature extraction is arranged in a fixed field order to form the 199-dimensional Token

$$\boldsymbol{t} = [t_1, t_2, \dots, t_D]^{\mathrm{T}}, \qquad D = 199.$$

The fields retain the same meanings and order and remain numerically finite during training, validation, and inference. Because the evidence fields have different scales and dynamic ranges, each field is standardized. Let the training set contain M samples. The mean and standard deviation of the jth field are

$$\mu_j = \frac{1}{M}\sum_{m=1}^{M} t_{m,j}, \qquad \sigma_j = \sqrt{\frac{1}{M}\sum_{m=1}^{M}(t_{m,j}-\mu_j)^2}.$$

The standardized model input is then

$$z_j = \mathrm{clip}\left(\frac{t_j-\mu_j}{\max(\sigma_j,\delta)}, -12,\ 12\right), \qquad j = 1, 2, \dots, D,$$

where $\delta = 10^{-6}$ and $\mathrm{clip}(u,a,b) = \min\{\max(u,a), b\}$. The means and standard deviations are determined only from the training samples of the corresponding network and remain fixed during validation, testing, and inference. Window-level root-mean-square normalization adjusts the root-mean-square amplitude of the observation sequence in each cognition window to a common reference, while field-wise standardization removes numerical-scale differences among evidence fields. The resulting $\boldsymbol{z}$ is the input to the cognition model.

2.2 Hierarchical Cognition Model Architecture

The Token is a fixed set of scalar evidence fields with defined meanings. It is not an equally spaced sequence and therefore has no intrinsic local translation structure. Considering cognition performance, robustness, model size, and hardware complexity, an MLP is used as the core of the cognition model. It directly combines all evidence through simple fixed-size matrix-vector operations. This structure makes the parameter counts, multiply-accumulate counts, and data paths known in advance and facilitates the assessment of implementation feasibility on an FPGA or DSP.

The processor's internal model is divided into four stages according to predefined cognition capabilities. The structure and parameters of each network are listed in Table 2.

Table 2. Hierarchical neural-network structure, parameter counts, and computational cost

Network	Input dimension	Hidden layers	Output dimension	Ensemble members	Parameters	Weight MACs per window
Stage 1	199	16	1	7	22,519	22,400
Stage 2	199	192, 96	7	5	288,035	286,560
Stage 3	206	192, 96	16	5	299,120	297,600
Stage 4	222	192, 96	6	5	309,630	308,160
Total	-	-	-	-	919,304	914,720

The 199-dimensional Token remains unchanged across all stages. The 206- and 222-dimensional inputs are formed by appending the 7-dimensional upstream support-probability vector and the combined 7+16-dimensional upstream support-probability vectors, respectively, to the original Token. Field-wise standardization does not change the dimension of a stage input. Stage 1 uses one hidden layer with a Gaussian error linear unit (GELU) activation. Stages 2-4 share the same two-hidden-layer structure, with hidden dimensions of 192 and 96. For stages 2-4, the forward transformation of one ensemble member is

$$\boldsymbol{h}_{\ell,1} = \mathrm{GELU}\left(\boldsymbol{W}_{\ell,1}\boldsymbol{z}_\ell + \boldsymbol{b}_{\ell,1}\right),$$

$$\boldsymbol{h}_{\ell,2} = \mathrm{GELU}\left(\boldsymbol{W}_{\ell,2}\boldsymbol{h}_{\ell,1} + \boldsymbol{b}_{\ell,2}\right), \qquad \boldsymbol{a}_\ell = \boldsymbol{W}_{\ell,3}\boldsymbol{h}_{\ell,2} + \boldsymbol{b}_{\ell,3},$$

where $\ell \in \{2, 3, 4\}$ denotes the network stage; $\boldsymbol{W}_{\ell,r}$ and $\boldsymbol{b}_{\ell,r}$ are the weight matrix and bias vector, respectively, of the rth linear transformation in stage ℓ, with $r \in \{1, 2, 3\}$; $\boldsymbol{h}_{\ell,1}$ and $\boldsymbol{h}_{\ell,2}$ are the outputs of the two hidden layers; $\boldsymbol{z}_\ell$ is the standardized input to the corresponding network; and $\boldsymbol{a}_\ell$ is the vector of uncalibrated logits.

Because the front-end Token represents signal structure from multiple physical views, the back end can use shallow networks to combine evidence and learn decision boundaries. In stages 2-4, the hidden dimensions decrease from 192 to 96 to limit the parameter count and pipeline depth. The architecture was not selected through an exhaustive search; it is an exploratory trade-off among cognition performance, training complexity, and engineering implementation.

The later stages do not perform single-path hard routing according to the maximum probability from an earlier stage. Instead, they use a soft-information cascade. Let $\boldsymbol{t} \in \mathbb{R}^{199}$ denote the complete Token, and let $\boldsymbol{p}_2 \in [0,1]^7$ and $\boldsymbol{p}_3 \in [0,1]^{16}$ denote the support-probability vectors from stages 2 and 3, respectively. The four network inputs are

$$\boldsymbol{u}_1 = \boldsymbol{t}, \qquad \boldsymbol{u}_2 = \boldsymbol{t},$$

$$\boldsymbol{u}_3 = [\boldsymbol{t}; \boldsymbol{p}_2] \in \mathbb{R}^{206}, \qquad \boldsymbol{u}_4 = [\boldsymbol{t}; \boldsymbol{p}_2; \boldsymbol{p}_3] \in \mathbb{R}^{222}.$$

The four networks are trained, calibrated, and frozen sequentially by stage, without joint end-to-end optimization. When stage 3 is trained, the frozen stage 2 produces $\boldsymbol{p}_2$. When stage 4 is trained, the frozen stages 2 and 3 produce $\boldsymbol{p}_2$ and $\boldsymbol{p}_3$. Gradients from later stages do not propagate into earlier stages. With the network architecture fixed, each stage independently optimizes its parameters, uses the validation set to select the training epoch to retain, and performs probability calibration.

Each stage uses fixed parameters determined from its own training samples to standardize $\boldsymbol{u}_\ell$ field by field. Because a later stage retains both the original Token and the complete upstream support-probability vectors, it does not discard alternative interpretations through a single discrete selection as a hard classification tree would.

To reduce variability due to a single training run or parameter initialization, each stage comprises multiple identically structured MLP members. Suppose stage ℓ contains E_ℓ members and the eth member outputs logit $a_{\ell,k}^{(e)}$ for class k. The ensemble output is first averaged in logit space:

$$\bar{a}_{\ell,k} = \frac{1}{E_\ell} \sum_{e=1}^{E_\ell} a_{\ell,k}^{(e)}.$$

Class-wise Platt scaling is then applied to map the averaged logits to support probabilities:

$$p_{\ell,k} = \sigma\left(\alpha_{\ell,k}\bar{a}_{\ell,k} + \beta_{\ell,k}\right), \qquad \sigma(v) = \frac{1}{1+\exp(-v)},$$

where $\alpha_{\ell,k} > 0$ and $\beta_{\ell,k}$ are determined from the designated calibration set and frozen together with the model parameters. The positive-slope constraint ensures that probability calibration does not reverse the ordering of the original scores. Stage 1 has only one logit; after ensemble averaging and calibration of the same form, it yields the signal-process probability p_{process} and the corresponding background probability $1 - p_{\text{process}}$.

Stages 2-4 use independent one-versus-rest (OvR) probabilistic models rather than mutually exclusive class-posterior models based on Softmax. Multiple semantics within a stage may have high support, or all semantics may have low support. If $p_{\ell,k}$ is the calibrated support probability of the current observation for semantic class k, then, in general,

$$\sum_{k=1}^{K_\ell} p_{\ell,k} \neq 1.$$

A strict hierarchical probabilistic model can be expressed as a conditional-probability chain:

$$P(Y_2, Y_3, Y_4 \mid \boldsymbol{t}) = P(Y_2 \mid \boldsymbol{t})P(Y_3 \mid Y_2, \boldsymbol{t})P(Y_4 \mid Y_2, Y_3, \boldsymbol{t}),$$

and node posterior probabilities can be obtained by multiplying probabilities along a path. The present

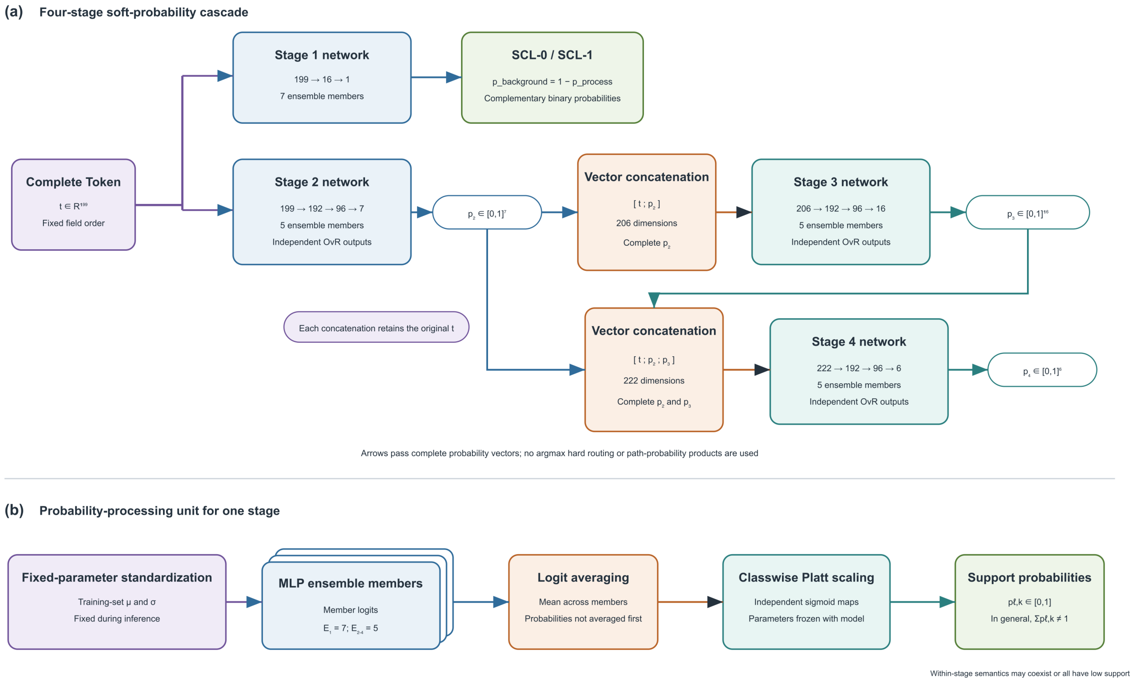

Figure 3: Hierarchical MLP soft-probability cascade and probability-processing unit for one stage

model retains only the coarse-to-fine direction of cognition-information transfer. Complete upstream support-probability vectors replace discrete parent nodes, and each later stage combines $\boldsymbol{t}$ with the upstream probabilities anew. This mechanism is related to classifier chains and stacked generalization [11, 12]. However, it neither multiplies probabilities along paths nor defines a joint distribution that satisfies parent-child consistency. The model is therefore a hierarchical soft cascade based on calibrated independent support probabilities rather than a strict Bayesian probability tree. Its outputs cannot be interpreted as marginal or conditional posteriors from a unified Bayesian model. The structure of the hierarchical soft cascade and the probability-processing unit within one stage are shown in Fig. 3.

2.3 Hierarchical Semantic Cognition and Sequential Output

At low SNRs, the multi-domain feature distributions of different signals may overlap substantially and become difficult to distinguish. The amount of evidence required also differs among cognition levels. If the evidence supports only a broader semantic level, forcing the model to output a fine-grained type conceals the uncertainty in the cognition result. The processor therefore represents signal cognition as a hierarchical probabilistic process that first distinguishes a signal process from the background and then estimates its basic form, signal family, and fine-grained type. The complete support-probability vector is retained at every stage.

The level of an output is denoted by the signal cognition level (SCL). The binary network in stage 1 corresponds to SCL-0 and SCL-1, while the other three networks correspond sequentially to SCL-2 through SCL-4. Their meanings are defined in Table 3.

Table 3. Signal cognition levels and output semantics

Cognition level	Cognition target	Output meaning
SCL-0	Reference background	Insufficient evidence in the current cognition window to support a signal process distinguishable from the reference background
SCL-1	Signal process	Observable signal activity distinguishable from the reference background is present in the current cognition window
SCL-2	Basic signal form	Support for seven basic forms: continuous single-carrier; multicarrier or multitone; spread-spectrum or pseudo-random-noise-coded; hopping or swept-frequency; coherent pulse or symbol burst; colored random process; and impulsive or burst random process
SCL-3	Signal family	Support for 16 signal families: narrowband tone; analog amplitude or sideband; analog frequency or angle; constant-envelope digital; amplitude-phase digital; spread-spectrum or pseudo-random-noise-coded; chirp or swept-frequency; frequency-hopping or stepped-frequency; tone or regular pulse; linear-frequency-modulated pulse; symbol-burst pulse; orthogonal-frequency-division-multiplexed or multitone; multicarrier; structured noise; impulsive noise; and burst noise
SCL-4	Fine-grained signal type	Support for five defined fine-grained types - constant-envelope digital, narrowband tone, orthogonal frequency-division multiplexing, phase-modulated digital, and amplitude-phase digital - together with an “other” state for signals outside these five types

As Table 3 shows, SCL-4 covers only five defined fine-grained types and assigns all remaining cases to “other.” The present model is therefore not a closed-set classifier for every modulation type. Further expansion of this level, or extension to SCL-5, would require new semantic definitions, training data, and classifiers.

Let the complete hierarchical probabilistic output for one cognition window be

$$\mathcal{C}(\boldsymbol{t}) = \left\{p_{\text{background}}, p_{\text{process}}, \boldsymbol{p}_2, \boldsymbol{p}_3, \boldsymbol{p}_4\right\},$$

where $p_{\text{background}} = 1 - p_{\text{process}}$, $\boldsymbol{p}_2 \in [0,1]^7$, $\boldsymbol{p}_3 \in [0,1]^{16}$, and $\boldsymbol{p}_4 \in [0,1]^6$. The sixth component of $\boldsymbol{p}_4$ denotes “other.” Thus, $\mathcal{C}(\boldsymbol{t})$ retains the support states of the same observation at different semantic levels rather than reducing them to one final label.

SCL-0 and SCL-1 use complementary background/process probabilities. Let p_{process} denote the signal-process probability, and let τ_0 and τ_1 be the background early-stop boundary and signal-process release boundary, respectively, where $0 \le \tau_0 < \tau_1 \le 1$. The corresponding action is

$$\mathcal{A}_{01} = \begin{cases} \text{background early stop,} & p_{\text{process}} \le \tau_0, \\ \text{defer the decision and retain soft cognition outputs,} & \tau_0 < p_{\text{process}} < \tau_1, \\ \text{signal-process release,} & p_{\text{process}} \ge \tau_1. \end{cases}$$

The boundaries τ_0 and τ_1 were determined in preliminary research and frozen before formal testing. Under “background early stop,” no later level is released. Following “signal-process release,” SCL-2 through SCL-4 can be interpreted. Between the two boundaries, the processor defers the background early-stop or signal-process release action while retaining the soft cognition outputs. Later networks do not alter p_{process} from stage 1 or the action determined from it.

SCL-2 through SCL-4 do not use a uniform hard acceptance threshold. The maximum OvR support is used only for diagnostic ranking and does not automatically constitute an exclusive class decision. Multiple high-support components represent compatible or competing interpretations, whereas uniformly low support indicates that the level lacks sufficient evidence for any known semantic class. The complete

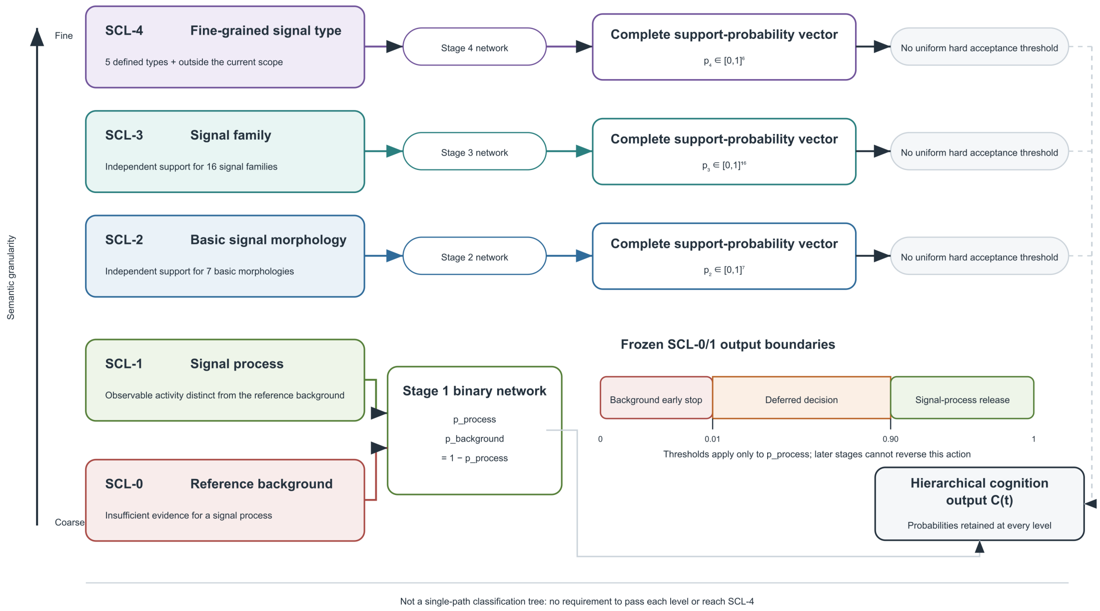

Figure 4: SCL semantics, support probabilities, and sequential outputs

vector therefore preserves states in which broader semantics are clear but fine-grained semantics remain uncertain, adjacent semantics coexist, or the observation belongs to the “other” state.

In the model, evidence from the same window is interpreted at successively more detailed semantic levels; its probability is not required to increase monotonically with either level or time. In addition, SCL-2 through SCL-4 share the original Token, and each later stage separately combines it with the complete upstream support-probability vectors. Consequently, a broader cognition result can be retained when fine-grained evidence is insufficient.

For a continuous I/Q stream, the model is applied repeatedly on a window-by-window basis, without cross-window recurrence or probability accumulation. SNR affects only the distinguishability of the evidence within a single window, while the window length determines the available observation span. The relationship among SCL semantics, support probabilities, and sequential outputs is shown in Fig. 4.

3 Simulation Results and Discussion

The simulations include single-carrier amplitude, phase, and frequency modulations; multicarrier and multitone signals; spread-spectrum and pseudo-random-noise-coded signals; and pulsed, burst, frequency-hopping, and swept-frequency forms, for a total of 73 specific signal types. SNR is defined as the ratio of signal power to noise power over the full sampled bandwidth within the cognition window and ranges from -12 to 16 dB.

The model is trained hierarchically. The training set contains 76,992 signal windows and 4,812 white-noise background windows. Formal testing uses the same simulation-generation framework as training, but the samples, random sequences, and observation perturbations are independent. The test set contains 3,079,680 signal windows and 192,480 background windows. The background early-stop and signal-process release boundaries used in formal testing are $\tau_0 = 0.01$ and $\tau_1 = 0.90$, respectively.

3.1 Overall Hierarchical Cognition Results

The binary background/signal-process network achieves an early-stop rate of 99.3558% on white-noise backgrounds, with 0.0010% of background windows incorrectly released as signal processes. For signal-containing windows at SNRs of 0 dB or higher, the signal-process release rate is 99.9479%. These results indicate that the binary network forms a relatively stable input boundary between the white-noise background and simulated signals defined in this study.

Table 4 reports the macro-averaged recall at each semantic level, together with the expected calibration error (ECE) and Brier score of the OvR probabilistic outputs.

Table 4. Recognition rate, correct-semantic support probability, and calibration metrics at each semantic level

Level	Macro-averaged recall (SNR $\geq$ 0 dB)	Mean correct-semantic support probability	OvR ECE	OvR Brier score
SCL-2	0.9240	0.9538	0.0642	0.0565
SCL-3	0.8763	0.9619	0.0784	0.0511
SCL-4	0.8715	0.8882	0.1662	0.1332

In Table 4, macro-averaged recall reflects the top-ranked semantic output, whereas the mean correct-semantic support probability is computed from the complete OvR vector. SCL-3 has higher correct-semantic support than SCL-2 but lower macro-averaged recall, indicating that compatible or competing semantics may be supported simultaneously. Passing only the maximum-probability result between stages would discard this uncertainty. The substantially higher ECE and Brier score at SCL-4 indicate that calibration of its fine-grained semantics is less accurate than at the preceding two levels.

3.2 Signal Cognition at Different Signal-to-Noise Ratios

Hussain et al. [2] evaluated a higher-order-cumulant classifier by recognition accuracy at different sample lengths and nominal SNRs under single-signal additive white Gaussian noise (AWGN) and Rayleigh channels. RadioML2018.01A has been used as input data in multiple studies [13, 14]. The dataset contains 24 digital and analog modulation classes, each sample comprises 1,024 complex points, and its SNR is labeled as E_s/N_0. DeepSig currently lists it as a historical dataset and provides a general notice that the early public datasets contain known errata [14]. On this dataset, Zhang et al. [5] reported closed-set classification accuracies of 80.22% at 4 dB and over 95% at 8 dB. Sig53 [15] defines SNR as E_s/N_0 in 4,096-sample narrowband single-signal windows and reports a best overall classification accuracy of 71.16%. WBSig53 [6], in contrast, retains the full input noise in wideband windows containing 1-6 signals and jointly evaluates detection, time-frequency localization, and modulation-family recognition using mean average precision (mAP) and mean average recall (mAR).

These studies use different observation conditions and evaluation conventions. This study considers fixed-length I/Q data without prior band localization; the true center frequency and true bandwidth are therefore not used for pre-inference filtering. Accordingly, SNR in the simulations is defined as the ratio of signal power to noise power over the full sampled bandwidth. Under this convention, simulations show that, at SNRs of 0 dB or higher, the process release rate is 99.9479%, while the macro-averaged recalls at the three semantic levels are 92.3971%, 87.6267%, and 87.1512%, respectively.

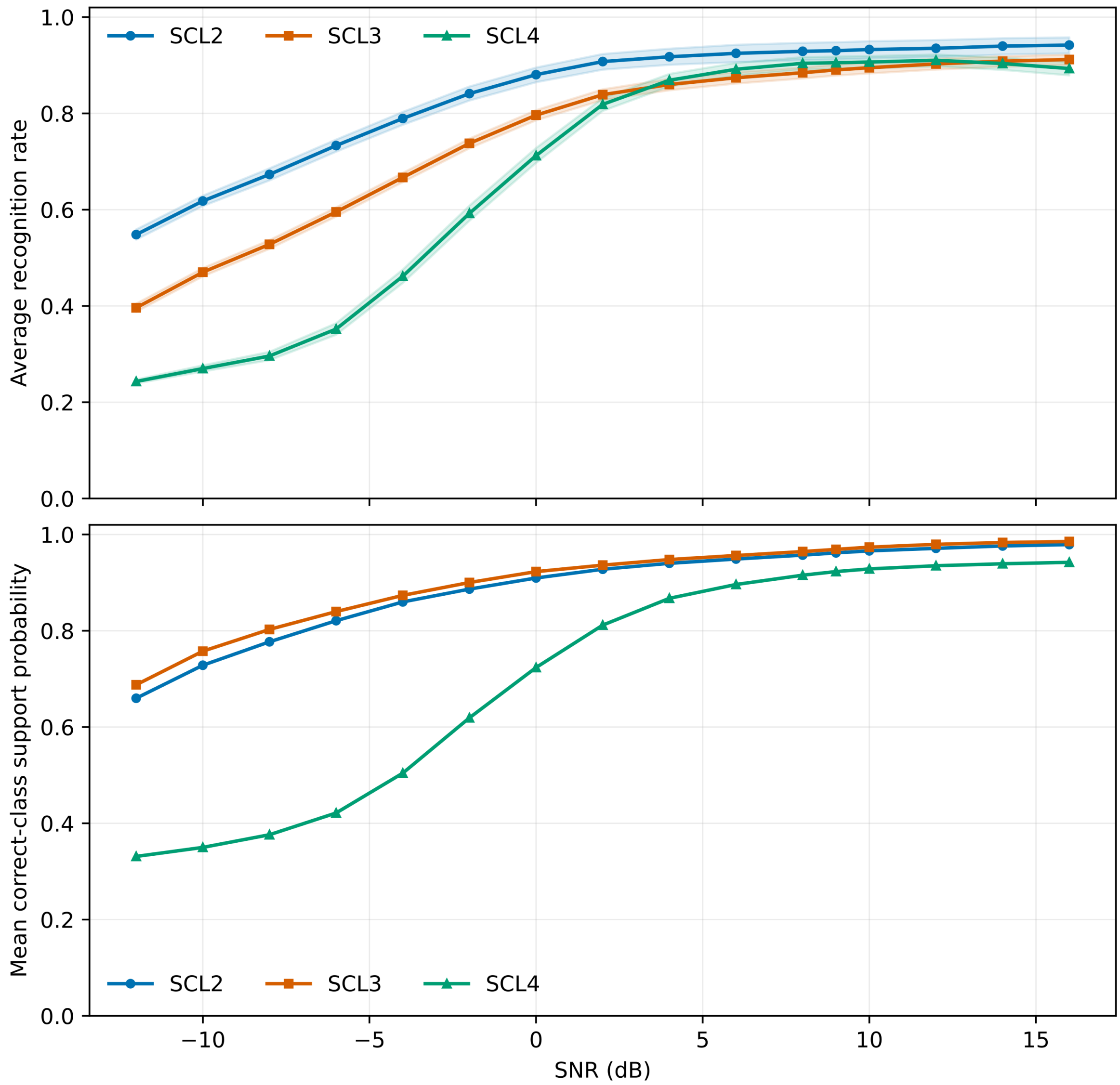

Figure 5: Hierarchical cognition results at different SNRs

Figure 5 shows the average recognition rate and correct-semantic support probability of the model at the three semantic levels as functions of SNR. As SNR increases from -12 dB to 0 dB, the average recognition rates of SCL-2, SCL-3, and SCL-4 increase from 54.82%, 39.62%, and 24.31% to 88.05%, 79.64%, and 71.24%, respectively. At 6 dB, they reach 92.48%, 87.41%, and 89.18%.

As Fig. 5 shows, the top-ranked semantic output does not improve strictly monotonically at higher SNRs. For phase-modulated digital signals, for example, the SCL-4 correct recognition rates at 12, 14, and 16 dB are 93.80%, 91.04%, and 86.63%, respectively, while the correct-semantic support probability remains close to 1. This phenomenon indicates that support strength, competing semantics, and top-ranked output at a fine-grained level are not equivalent: strengthening support for the correct semantic class does not necessarily eliminate competition from adjacent semantic classes.

Figure 6 presents hierarchical cognition heatmaps under different signal-class and SNR conditions. Pulsed and burst signals exhibit relatively stable support for their basic forms at lower SNRs, whereas digital modulation, multicarrier, and some noise classes generally require stronger evidence. Colored random processes, structured noise, and narrowband tones are the most difficult classes for the current model. These differences indicate that the level of detail that can be reached is jointly constrained by SNR, the structures observable within a finite window, and overlap between the evidence for adjacent semantics.

3.3 Validation with an Independent Signal Source

To examine whether the test results depend on the simulated-signal source, scikit-dsp-comm [16] is used as an independent third-party signal generator. Validation is conducted on quadrature phase-shift keying (QPSK), 16-ary quadrature amplitude modulation (16-QAM), and 64-ary quadrature amplitude modulation (64-QAM), which are supported by both scikit-dsp-comm and TorchSig [17]. The two sources independently generate continuous I/Q waveforms with matched sampling rates, samples per symbol, center-frequency offsets, SNRs, and energy bandwidths. Twenty waveform pairs are generated for each specified physical condition, yielding 4,320 pairs of input samples. The results at SNRs of 0 dB or higher are listed in Table 5.

Table 5. Comparison of cognition results for TorchSig and the independent signal source (SNR $\geq$ 0 dB)

Level	TorchSig correct recognition rate	Independent-source correct recognition rate	TorchSig correct-semantic support probability	Independent-source correct-semantic support probability
SCL-2	0.8222	0.7770	0.8936	0.8936
SCL-3	0.8570	0.8674	0.9606	0.9665
SCL-4	0.9600	0.9704	0.9395	0.9440

Under these matched physical conditions, the third-party and TorchSig signals show good agreement in the support probabilities for the correct semantics at all levels and in the SCL-3 and SCL-4 cognition results.

4 Model Size and Engineering Feasibility Assessment

Model computation consists of multi-view feature extraction and hierarchical neural-network inference. The former includes finite-length operators for spectra, instantaneous frequency, correlation, statistical summaries, and time-frequency trajectories. The latter consists of fixed-size matrix-vector operations, activation functions, and probability calibration. Its data flow and computational cycles can be determined in advance.

4.1 Model Size and Computational Cost

The hierarchical neural network contains 919,304 trainable parameters. With 16-bit fixed-point weights, the theoretical storage requirement is approximately 1.84 MB. Completing one cognition pass through the four stages requires approximately 914,720 weight multiply-accumulate operations, or 0.915 million multiply-accumulate operations (MMAC). Because each stage uses a relatively small MLP, the back end can reuse a shared matrix-computation unit sequentially across stages rather than requiring four independent compute arrays to operate simultaneously.

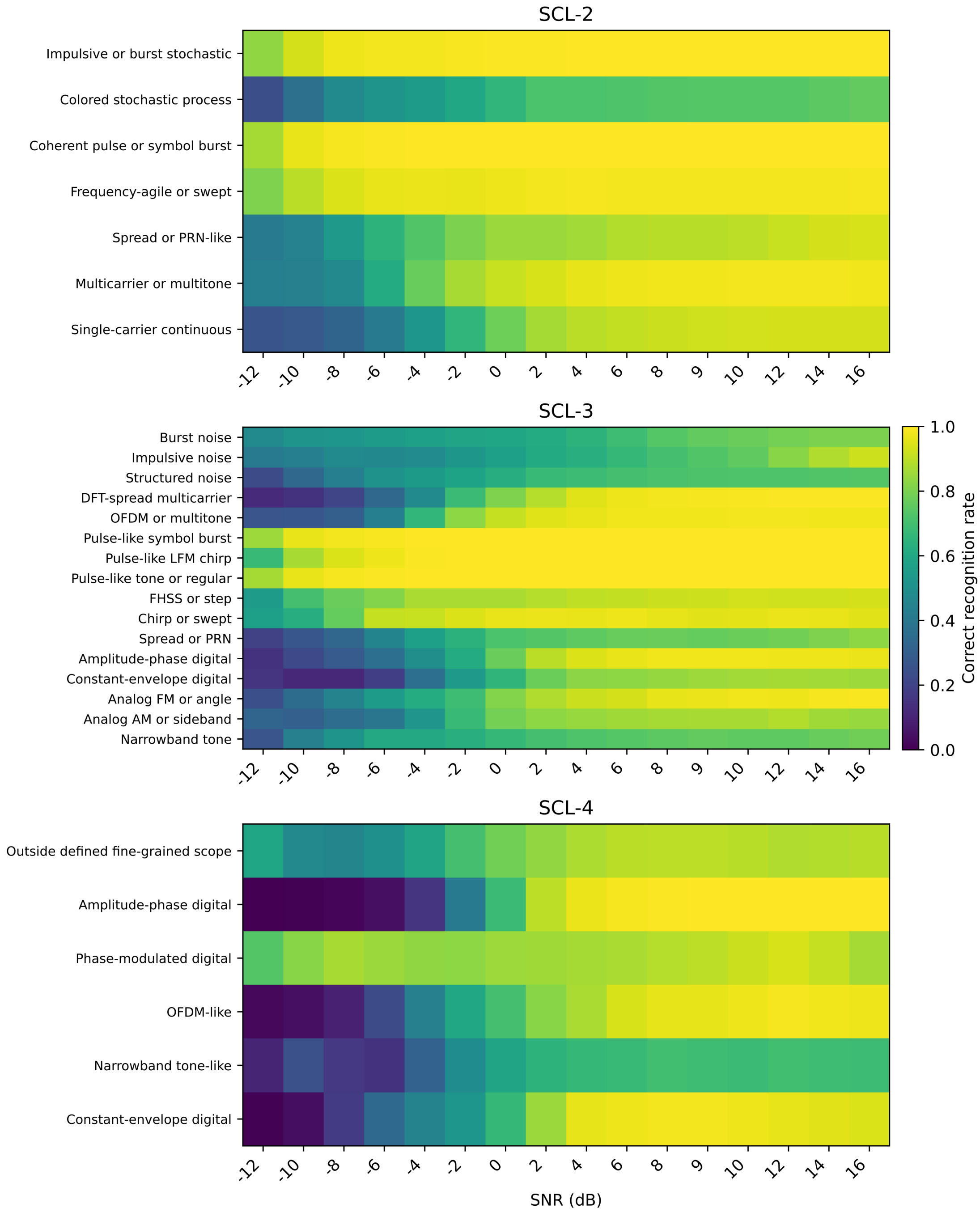

Figure 6: Hierarchical cognition heatmaps across signal classes and SNRs

The feature-extraction front end can be organized into four pipelined units: a shared input buffer, spectral and time-frequency transforms, correlation and statistical summaries, and feature output. Normalized views, spectral frames, and some intermediate statistics can be reused across multiple features. Streaming-equivalence tests show that repeated normalization of the complete observation can be reduced while preserving the 199-dimensional Token and model outputs. Means, energies, finite-delay correlations, and fixed-length Fourier transforms can be implemented directly with fixed-point pipelines. Quantiles, entropy, logarithms, and nonlinear activation functions require histogram-based, lookup-table, or piecewise approximations. Thus, the main hardware-mapping issues are not whether the neural-network parameters can be stored, but how to control intermediate front-end bit widths, schedule shared transform units, and manage feature deviations caused by approximate operators. Under the current design, each cognition window contains 727 mutually independent framewise spectral transforms, which can be grouped into 12 batches according to their lengths and data layouts. For non-overlapping 8192-sample cognition windows, the estimated number of windows per second is proportional to the input sampling rate.

For engineering estimates, suppose the input is represented by 16 bits for each of the I and Q channels. Let the sampling rate be expressed in megasamples per second (MS/s), the input data rate in gigabytes per second (GB/s), and the back-end computational requirement in giga multiply-accumulate operations per second (GMAC/s). Table 6 lists the input data rate, window update rate, and current back-end computational requirement at several input sampling rates.

Table 6. Data rates and back-end computational requirements at different input sampling rates

Sampling rate	Input data rate	Time per window	Windows/s	Back-end requirement
200 MS/s	0.8 GB/s	40.960 μs	24,414	22.3 GMAC/s
250 MS/s	1.0 GB/s	32.768 μs	30,518	27.9 GMAC/s
500 MS/s	2.0 GB/s	16.384 μs	61,035	55.8 GMAC/s

In computer simulations, the 199-dimensional feature vector uses 32-bit floating-point values and occupies 796 bytes. If each field is affine-quantized to 16 bits using a range determined from the training data, the vector size is reduced to 398 bytes. In the quantization validation, the worst change in macro-averaged recall is -0.000003, and the maximum observed decrease under different signal and SNR conditions does not exceed 0.001225. Eight-bit quantization likewise causes only small changes in the aggregate metrics, but non-monotonic errors may occur in some small-sample conditions. Sixteen bits is therefore a more conservative width for the feature interface. If the 8192-sample window length is changed according to actual signal conditions, the corresponding feature scales and model parameters must be retrained and revalidated.

4.2 Feasibility Boundaries

In terms of parameter storage, back-end multiply-accumulate scale, finite-window operators, and streaming data dependencies, the proposed cognition model has the basic structure needed for mapping to a single FPGA or a cooperative FPGA-DSP platform. For non-overlapping window processing at 200-250 MS/s, the back-end computation can be implemented using parallel multiply-accumulate arrays and hierarchical reuse. At 500 MS/s, the back-end workload remains finite and parallelizable, but front-end data movement, time-frequency transforms, and statistical summaries require an optimized design with greater parallelism or channelization.

5 Conclusion

This paper examined a processing method for multi-class signal cognition in continuous I/Q streams based on multi-view physical evidence, a hierarchical neural network, and a soft probabilistic cascade. The model extracts features from the time, frequency, and time-frequency domains and at different time scales within fixed-length windows. It then produces probabilistic cognition results for signal process, basic signal form, signal family, and fine-grained type in sequence. Unlike a classifier that outputs only one class, this architecture retains states at different semantic levels and preserves broader cognition information when fine-grained evidence is insufficient or adjacent semantics coexist.

Simulation results show that the model produces cognition results from broader to finer levels across multiple signal types and wide ranges of SNR and occupied bandwidth. The average recognition rate at each level generally increases with SNR. The assessment of model complexity and computational cost indicates that the model has the basic structure needed for mapping to FPGA or DSP platforms. The multi-view observation and hierarchical probabilistic-output method developed in this study organizes cognition information at different levels within a single processing chain and provides an approach to general-purpose cognition of complex signals.

References

[1] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su. Survey of automatic modulation classification techniques: Classical approaches and new trends. *IET Communications*, 1(2):137–156, 2007. doi: 10.1049/iet-com:20050176. URL `https://doi.org/10.1049/iet-com:20050176`.

[2] Asad Hussain, Sheraz Alam, Sajjad A. Ghauri, Mubashir Ali, Husnain Raza Sherazi, Adnan Akhunzada, Iram Bibi, and Abdullah Gani. Automatic modulation recognition based on the optimized linear combination of higher-order cumulants. *Sensors*, 22(19):7488, 2022. doi: 10.3390/s22197488. URL `https://doi.org/10.3390/s22197488`.

[3] Timothy J. O'Shea and Jakob Hoydis. An introduction to deep learning for the physical layer, 2017. URL `https://arxiv.org/abs/1702.00832`.

[4] Zhiyu Qu, Xiaojie Mao, and Zhian Deng. Radar signal intra-pulse modulation recognition based on convolutional neural network. *IEEE Access*, 6:43874–43884, 2018. doi: 10.1109/ACCESS.2018.2864347. URL `https://doi.org/10.1109/ACCESS.2018.2864347`.

[5] Xinliang Zhang, Tianyun Li, Pei Gong, Renwei Liu, and Xiong Zha. Modulation recognition of communication signals based on multimodal feature fusion. *Sensors*, 22(17):6539, 2022. doi: 10.3390/s22176539. URL `https://doi.org/10.3390/s22176539`.

[6] Luke Boegner, Garrett Vanhoy, Phillip Vallance, Manbir Gulati, Dresden Feitzinger, Bradley Comar, and Robert D. Miller. Large scale radio frequency wideband signal detection & recognition, 2022. URL `https://arxiv.org/abs/2211.10335`.

[7] Yi Shi, Kemal Davaslioglu, Yalin E. Sagduyu, William C. Headley, Michael Fowler, and Gilbert Green. Deep learning for RF signal classification in unknown and dynamic spectrum environments, 2019. URL `https://arxiv.org/abs/1909.11800`.

[8] Huiying Niu, Xun Xie, Xiaojing Cheng, and Jing Bai. Open-set automatic modulation recognition based on circular prototype learning and denoising diffusion model. *Electronics*, 14(3):430, 2025. doi: 10.3390/electronics14030430. URL `https://doi.org/10.3390/electronics14030430`.

[9] Elsayed Mohammed, Omar Mashaal, and Hatem Abou-Zeid. Using early exits for fast inference in automatic modulation classification. In *2023 IEEE Global Communications Conference (GLOBECOM)*, 2023. doi: 10.1109/GLOBECOM54140.2023.10436917. URL `https://arxiv.org/abs/2308.11100`.

[10] Shani Goren, Ido Galil, and Ran El-Yaniv. Hierarchical selective classification. In *Advances in Neural Information Processing Systems*, volume 37, 2024. doi: 10.52202/079017-3526. URL `https://proceedings.neurips.cc/paper_files/paper/2024/hash/c8b100b376a7b338c84801b699935098-Abstract-Conference.html`.

[11] Jesse Read, Bernhard Pfahringer, Geoff Holmes, and Eibe Frank. Classifier chains for multi-label classification. *Machine Learning*, 85(3):333–359, 2011. doi: 10.1007/s10994-011-5256-5. URL `https://doi.org/10.1007/s10994-011-5256-5`.

[12] David H. Wolpert. Stacked generalization. *Neural Networks*, 5(2):241–259, 1992. doi: 10.1016/S0893-6080(05)80023-1. URL `https://doi.org/10.1016/S0893-6080(05)80023-1`.

[13] Timothy J. O'Shea, Tamoghna Roy, and T. Charles Clancy. Over-the-air deep learning based radio signal classification. *IEEE Journal of Selected Topics in Signal Processing*, 12(1):168–179, 2018. doi: 10.1109/JSTSP.2018.2797022. URL `https://doi.org/10.1109/JSTSP.2018.2797022`.

[14] DeepSig. Datasets: Historical dataset—RADIOML 2018.01a. Web page. URL `https://www.deepsig.ai/datasets/`. Accessed 2026-07-31.

[15] Luke Boegner, Manbir Gulati, Garrett Vanhoy, Phillip Vallance, Bradley Comar, Silvija Kokalj-Filipovic, Craig Lennon, and Robert D. Miller. Large scale radio frequency signal classification, 2022. URL `https://arxiv.org/abs/2207.09918`.

[16] Mark Wickert. scikit-dsp-comm: DSP and communications package. Software, version 2.1.0. URL `https://github.com/mwickert/scikit-dsp-comm`. Accessed 2026-07-29.

[17] TorchSig Team. TorchSig: Signal processing machine learning toolkit. Software, version 2.1.1. URL `https://github.com/TorchDSP/torchsig`. Accessed 2026-07-29.