

# Joint Recognition of LPI Radar Signals Using a VLM with TFD-Text Alignment

A Vision-Language Approach for Waveform Classification and Parameter Estimation of LPI Radars

Jaehyeok Yoon  
Department of Electrical and Electronic  
Engineering  
Hanyang University  
Ansan, South Korea  
[serp82@hanyang.ac.kr](mailto:serp82@hanyang.ac.kr)

Haewoon Nam\*  
Department of Electrical and Electronic  
Engineering  
Hanyang University  
Ansan, South Korea  
[hnam@hanyang.ac.kr](mailto:hnam@hanyang.ac.kr)

Jaerock Kwon  
Department of Electrical and Computer  
Engineering  
University of Michigan-Dearborn  
Dearborn, MI, USA  
[jrkwon@umich.edu](mailto:jrkwon@umich.edu)

**Abstract**— This paper proposes a vision-language model for the joint recognition of Low Probability of Intercept (LPI) radar signals through time-frequency distribution (TFD)-text alignment. The proposed framework unifies waveform classification and signal parameter estimation by aligning TFD spectrograms with hierarchical textual prompts in a shared embedding space. To support both general waveform type recognition and fine-grained parameter inference, we introduce a prompt dropout strategy that balances rich and simple prompts during training. Evaluated on multiple TFD representations including SPWVD, CWD, and SAFI, the model demonstrates high accuracy and interpretability across both tasks. This unified approach offers a compact, extensible solution for LPI radar signal understanding.

**Keywords**— *Vision-Language Model, LPI Radar Signal Recognition, Time-Frequency image, Waveform Classification*

## I. INTRODUCTION

Effective analysis of radar signals in electronic intelligence and communications intelligence scenarios requires answering two fundamental questions: (1) what type of signal is present (waveform classification) [1, 2], and (2) how the signal was generated (parameter estimation) [3, 4]. These two tasks are essential for interpreting modern Low Probability of Intercept (LPI) radar systems, which intentionally obscure their signal characteristics through various modulation techniques.

Conventional deep learning approaches typically decouple these tasks. A convolutional neural network (CNN) might be trained on time-frequency distribution (TFD) images to classify signal types [5], such as LFM, Frank, or Barker. Once classified, a separate model is used to estimate relevant parameters, such as sweep direction or code length. This two-stage architecture leads to multiple limitations: it requires specialized models per waveform type, lacks generalization across signal families, and results in increased complexity during deployment.

To address these challenges, we propose a vision-language model (VLM) for joint recognition of LPI radar signals. Our method treats the TFD spectrogram as a visual input and the associated waveform name and parameters as a structured textual prompt. The model learns to align these two modalities within a shared embedding space, enabling simultaneous

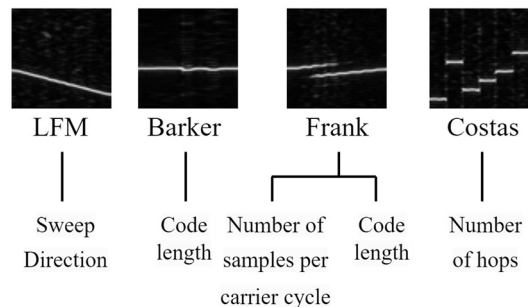


Fig. 1. Examples of TFD for the four LPI radar waveforms used in this paper. Each waveform is associated with a characteristic modulation parameter: sweep direction (LFM), code length (Barker), Number of samples per carrier cycle and code length (Frank), and number of hops (Costas).

learning of signal identity and modulation characteristics. Fig. 1 summarizes the waveform types and parameters that the VLM is trained to associate, illustrating the diversity of TFD signatures considered in this paper.

## II. RELATED WORKS

TFD techniques such as the Choi-Williams Distribution (CWD), Smoothed Pseudo-Wigner-Ville Distribution (SPWVD), and Short-Time Fourier Transform (STFT) have long been employed in radar signal analysis, particularly for non-stationary and LPI waveforms. Deep learning approaches leveraging TFD images have demonstrated success in classifying radar signal types using CNNs [5, 6, 7], but these models typically address only the waveform classification task. Parameter estimation—such as determining the sweep direction of an LFM signal or the code length of a Barker sequence—has conventionally been handled separately, often using rule-based post-processing, regression modules, or task-specific classifiers [8]. This separation introduces challenges in scalability, pipeline complexity, and cumulative error.

Recent advances in VLMs, most notably CLIP [9], have shown that joint embedding spaces for image-text pairs can enable both discriminative recognition and semantic reasoning. These models, trained via contrastive learning, excel at aligning visual content with descriptive language and have been widely applied in domains such as remote sensing, medical imaging, and document understanding. However, their application to radar signal intelligence remains largely unexplored, particularly in the context of structured signal parameters embedded within spectrograms.

\* Corresponding author

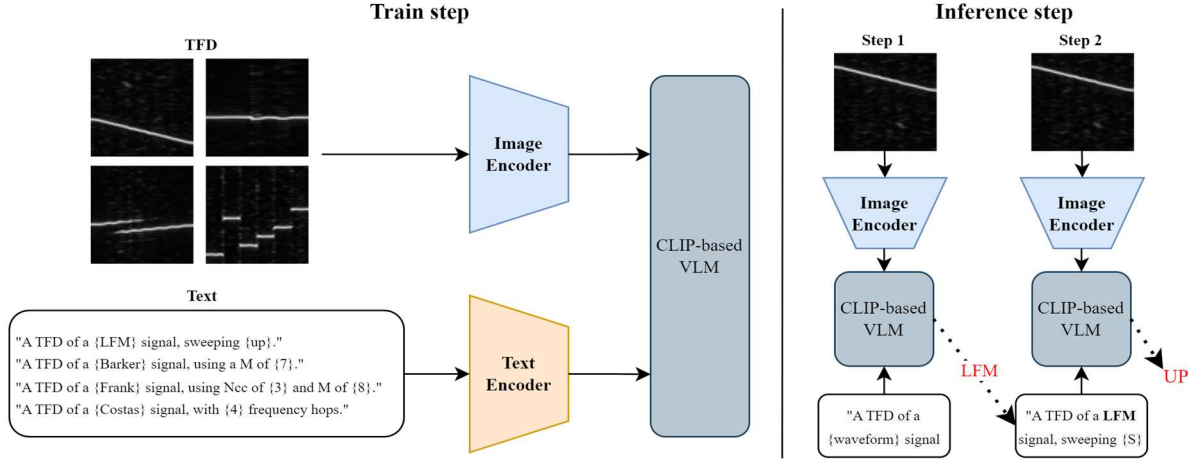


Fig. 2. Training and inference workflow of the proposed TFD-text aligned vision-language model. During training (left), TFD images and hierarchical textual prompts are jointly encoded by the image and text encoders within a CLIP-based VLM to learn a shared embedding space. During inference (right), the model performs two sequential steps: **Step 1** predicts the waveform type using a simple prompt, and **Step 2** predicts the corresponding modulation parameter using a refined rich prompt conditioned on the Step-1 result (e.g., sweep direction for LFM). This two-stage inference enables joint waveform classification and parameter estimation within a single VLM classification and parameter inference through language-guided learning

In related multimodal work, hierarchical prompts and compositional reasoning have been used to guide models toward fine-grained understanding. While similar ideas have been explored in natural image domains [10], signal processing tasks have only recently begun to adopt such techniques [11].

Our work situates itself at the intersection of these research areas by leveraging TFD-text alignment within a VLM framework to perform joint recognition of radar waveform types and their modulation parameters. Unlike previous methods that require separate pipelines, our approach enables integrated, scalable, and interpretable radar signal understanding through a single unified architecture.

### III. METHOD

This section describes the proposed method for jointly recognizing LPI radar waveforms and estimating their signal parameters using a unified VLM. The overall framework is built on a contrastive learning architecture that aligns TFD images with structured textual prompts, as illustrated in Fig. 2. Unlike conventional approaches that separate classification and parameter estimation into distinct modules, our method uses a single model to perform both tasks simultaneously by learning a shared multimodal embedding space.

#### A. Contrastive Vision-Language Modeling on TFD

We adopt a CLIP-based VLM consisting of two parallel encoders: a visual encoder that processes TFD images, and a text encoder that processes natural language prompts describing the corresponding radar signal. Each TFD image and its associated textual description form a positive pair, while other TFD-text combinations are treated as negatives. During training, the model learns to project both modalities into a shared embedding space, such that the cosine similarity between a matched TFD-text pair is maximized, and mismatched pairs are minimized [9].

Formally, given a batch of  $N$  TFD-text pairs  $\{(x_i, t_i)\}_{i=1}^N$ , where  $x_i$  is the  $i$ -th the TFD image and  $t_i$  is its corresponding  $i$ -th text prompt. And the model computes the similarity scores  $s_{ij}$ :

$$s_{ij} = \cos(f_{TFD}(x_i), f_{text}(t_j)), \quad (1)$$

where  $f_{TFD}$  and  $f_{text}$  are the TFD and text encoders, respectively. The loss is computed as a symmetric cross-entropy over the TFD-to-text and text-to-TFD matching logits, following the original CLIP formulation. This contrastive training enables the model to associate signal semantics, including waveform type and generation parameters, directly from the TFD-text alignment.

#### B. Hierarchical Text Prompt

To guide the model in learning both waveform type and parameter semantics, we construct two levels of textual prompts for each signal:

- **Simple prompts** are short descriptions that only specify the waveform type, such as: "A TFD of a LFM signal.", "A TFD of a Frank signal."
- **Rich prompts** include additional details about signal parameters, such as sweep direction, code length, or hop count. Examples include: "A TFD of a LFM signal, sweeping down.", "A TFD of a Frank signal, using the number of samples per carrier cycle ( $N_{cc}$ ) of 3 and  $M$  of 8.", "A TFD of a Costas signal, with 4 frequency hops."

These prompts are generated programmatically using metadata from the synthetic radar signals. The use of rich prompts allows the model to learn signal-specific characteristics beyond class labels, forming the basis for parameter inference through vision-language alignment.

#### C. Prompt Dropout Strategy

During training, a key challenge arises when the model is exposed only to rich prompts: it may overfit to small token differences (e.g., " $N_{cc}$  of 3" vs. " $N_{cc}$  of 4") and ignore higher-level concepts shared across classes. As a result, it may fail to generalize when presented with simple prompts during inference. To address this issue, we propose a prompt dropout mechanism.

Prompt Dropout randomly replaces a subset of rich prompts with simple prompts at a fixed ratio during training

TABLE I. PARAMETER OF RADAR WAVEFORMS

Type of waveform	Parameter	Values
LFM	Sweep Direction	{Up, Down}
Barker	Code length	{2, 3, 4, 5, 7, 11, 13}
Frank	Number of samples per carrier cycle	{3, 4, 5}
	Code length	{6, 7, 8}
Costas	Number of hops	{3, 4, 5, 6}

(e.g., 20%). This forces the model to attend to both the global waveform identity and the discriminative parameter details. Specifically, for each training batch, we sample whether to use the rich or simple prompt based on the simple prompt ratio parameter. This creates a diverse training signal that encourages the model to capture hierarchical semantic structure.

#### D. Validation Strategy

The validation strategy is designed to separately evaluate the two core tasks of the model: Waveform classification (Step 1) is evaluated using simple prompts in a 1:N retrieval setup, where the model must match each TFD image to the correct waveform prompt out of all possible candidates.

Parameter estimation (Step 2) is evaluated using a custom inspection tool rather than through direct validation loss. This tool samples predictions from the trained model using rich prompts, then compares extracted parameter values (e.g., LFM sweep direction, Frank's  $N_{cc}/M$  values) with the ground truth. By decoupling the evaluation of classification and parameter inference, we can better isolate the model's strengths and limitations across hierarchical reasoning tasks.

## IV. EXPERIMENTS

We evaluate the proposed VLM-based framework on two core tasks: waveform classification (Step 1) and parameter estimation (Step 2). Experiments were conducted using synthetically generated LPI radar signals and three types of TFD representations: SPWVD, CWD, and SAFI [13]. All results are reported on held-out validation sets with no overlap between training and test samples. For each waveform type, 1,000 samples were generated for multiple SNR conditions, and the corresponding modulation parameters were assigned according to the configurations summarized in TABLE I. The dataset was split into training and validation sets using an 8:2 ratio, ensuring that all reported results are measured on held-out samples with no overlap between training and test data.

#### A. Waveform Classification (Step 1)

Waveform classification performance was further evaluated across varying SNR conditions, as shown in Fig. 3. All three TFD representations—CWD, SAFI, and SPWVD—exhibited a consistent upward trend in accuracy as SNR increased. At low SNR values (e.g., -16 dB), SAFI provided the highest robustness with an accuracy above 0.85, while CWD and SPWVD showed lower performance due to increased noise artifacts in their TFD patterns. As the

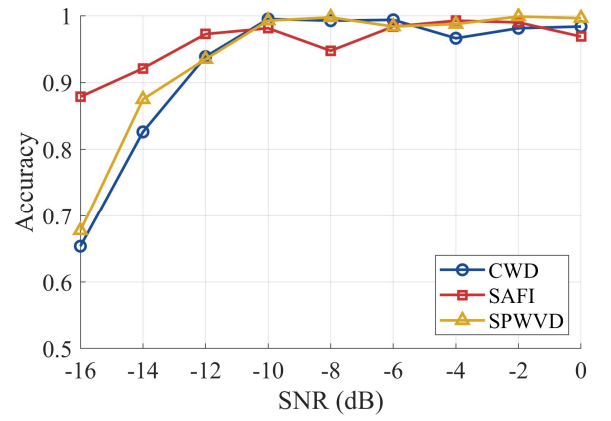


Fig. 3. Step-1 waveform classification accuracy vs. SNR

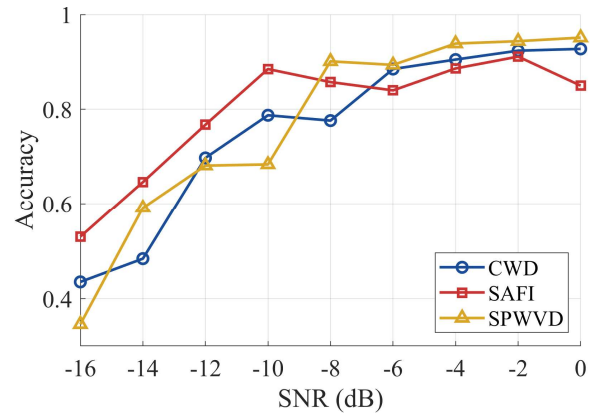


Fig. 4. Step-2 overall parameter estimation accuracy vs. SNR.

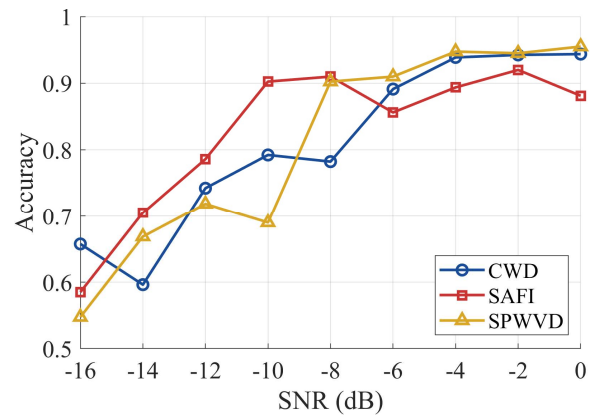


Fig. 5. Full pipeline accuracy combining Step-1 and Step-2 outcomes.

SNR improved beyond -12 dB, all three methods converged to high classification accuracy, reaching above 0.95 at -10 dB and achieving near-perfect accuracy ( $\approx 1.0$ ) from -8 dB to 0 dB. Across the entire SNR range, SPWVD and SAFI demonstrated slightly higher stability than CWD, but differences remained modest.

#### B. Parameter Estimation (Step 2)

Parameter estimation accuracy was evaluated across different SNR levels, as shown in Fig. 4. All three TFD representations—CWD, SAFI, and SPWVD—show a clear improvement in accuracy as the SNR increases. At very low SNR values (e.g., -16 dB to -12 dB), SAFI exhibits the

strongest robustness with noticeably higher accuracy, while CWD and SPWVD show more sensitivity to noise. As the SNR approaches  $-10$  dB, all three methods begin to converge, with SAFI consistently maintaining a performance advantage.

Between  $-8$  dB and  $-4$  dB, SPWVD and SAFI both reach high accuracy levels ( $\approx 0.9$  or higher), indicating that their TFD structures preserve parameter-relevant features even under moderate noise conditions. At SNR values from  $-6$  dB upward, SPWVD slightly outperforms the other two methods, achieving near-saturation accuracy around  $0.95$ – $1.0$ , while CWD and SAFI maintain strong but slightly lower performance. These trends demonstrate that the proposed VLM can reliably infer fine-grained waveform parameters across a wide range of SNR conditions, with particularly stable performance when SPWVD or SAFI representations are used.

### C. Full Pipeline Accuracy (Step 1, 2)

The full pipeline performance, obtained by jointly evaluating waveform classification and parameter estimation, is shown in Fig. 5. As SNR increases, the overall accuracy improves steadily across all three TFD representations. At very low SNR levels (below  $-12$  dB), SAFI demonstrates the highest robustness, maintaining noticeably better accuracy compared to CWD and SPWVD. Around the mid-SNR region ( $-10$  dB to  $-8$  dB), the three methods begin to converge, with SAFI still showing a slight advantage.

However, once the SNR exceeds  $-8$  dB, SPWVD consistently achieves the highest full-pipeline accuracy, approaching saturation above  $0.95$  as the noise level decreases. These results indicate that SAFI is more resilient in severely noisy conditions ( $\text{SNR} < -8$  dB), while SPWVD provides superior performance when the SNR is greater than  $-8$  dB, making it the most reliable representation in moderate-to-high SNR environments.

## V. CONCLUSION

This paper presented a unified vision–language framework for joint waveform classification and parameter estimation of LPI radar signals using TFD–text alignment. By leveraging a CLIP-style contrastive learning architecture, the proposed model learns a shared embedding space that effectively integrates TFD-based visual information with structured textual descriptions. The introduction of hierarchical prompts and the prompt dropout strategy proved essential in enabling the model to generalize across varying levels of textual detail, supporting both coarse waveform identification and fine-grained parameter inference within a single model.

Extensive experiments across three TFD representations demonstrated the effectiveness of the proposed approach. Step-1 waveform classification and Step-2 parameter estimation both achieved high accuracy, and the combined full-pipeline evaluation further confirmed the robustness of the model under noisy conditions. Notably, SNR-dependent performance analysis revealed a clear trend: SAFI provides the highest accuracy in low-SNR environments ( $\text{SNR} < -8$  dB), while SPWVD offers superior performance in medium-to-high SNR regions ( $\text{SNR} \geq -8$  dB). These results highlight the adaptability of the multimodal

framework to different TFD structures and noise characteristics.

Overall, the proposed VLM-based architecture offers a compact, scalable, and interpretable solution for radar signal understanding, replacing conventional multi-stage pipelines with a single, unified model. Future work will extend the framework to real-world radar datasets, incorporate more diverse modulation schemes, and explore question-driven VQA formulations to enable interactive and task-aware signal intelligence.

## ACKNOWLEDGMENT

This research was supported by the MSIT (Ministry of Science, ICT), Korea, under the Global Research Support Program in the Digital Field program (RS-2024-00428465) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation)

## REFERENCES

- [1] KISHORE, Thokala Ravi; RAO, K. Deerga. Automatic intrapulse modulation classification of advanced LPI radar waveforms. *IEEE Transactions on Aerospace and Electronic Systems*, 2017, 53.2: 901-914.
- [2] VANHOY, Garrett; SCHUCKER, Thomas; BOSE, Tamal. Classification of LPI radar signals using spectral correlation and support vector machines. *Analog integrated circuits and signal processing*, 2017, 91.2: 305-313.
- [3] TAO, Wan, et al. Research on LPI radar signal detection and parameter estimation technology. *Journal of Systems Engineering and Electronics*, 2021, 32.3: 566-572.
- [4] CHILUKURI, Raja Kumari; KAKARLA, Hari Kishore; SUBBARAO, K. Estimation of modulation parameters of LPI radar using cyclostationary method. *Sensing and Imaging*, 2020, 21.1: 51.
- [5] KONG, Seung-Hyun, et al. Automatic LPI radar waveform recognition using CNN. *Ieee Access*, 2018, 6: 4207-4219.
- [6] GUO, Qiang; YU, Xin; RUAN, Guoqing. LPI radar waveform recognition based on deep convolutional neural network transfer learning. *Symmetry*, 2019, 11.4: 540.
- [7] MA, Zhiyuan, et al. LPI radar waveform recognition based on features from multiple images. *Sensors*, 2020, 20.2: 526.
- [8] KUMARI, Chilukuri Raja; KAKARLA, Hari Kishore; SUBBARAO, K. Estimation of intrapulse modulation parameters of LPI radar under noisy conditions. *International Journal of Microwave and Wireless Technologies*, 2022, 14.9: 1177-1194.
- [9] RADFORD, Alec, et al. Learning transferable visual models from natural language supervision. In: *International conference on machine learning*. PmlR, 2021. p. 8748-8763.
- [10] WANG, Henan, et al. Hierarchical Prompt Learning for Compositional Zero-Shot Recognition. In: *IJCAI*. 2023. p. 3.
- [11] ZHAO, Yurui, et al. Zero-shot Automatic Modulation Recognition Using a Large Vision-Language Model. *IEEE Transactions on Communications*, 2025.
- [12] FENG, Hancong; JIANG, KaiLI. NON-COOPERATIVE RADAR SIGNAL PARSING. *arXiv preprint arXiv:2503.15213*, 2025.
- [13] LIN, Anni, et al. Unknown radar waveform recognition based on transferred deep learning. *IEEE Access*, 2020, 8: 184793-184807