A UNIFIED RF FINGERPRINTING FRAMEWORK FOR DEVICE LEVEL UAV IDENTIFICATION USING HOCv2, WAVELET RF-DNA, AND DEEP LEARNING

Nguyen Nhu Y*, Tran Cong Trang

Naval Academy, Vietnam

*Corresponding author: nguyeny.hvhq@gmail.com

(Received: March 26, 2026; Revised: May 22, 2026; Accepted: May 25, 2026)

DOI: 10.31130/ud-jst.2026.24(6A).190E

Abstract - The rapid proliferation of unmanned aerial vehicles (UAVs) introduces significant challenges for spectrum surveillance and airspace security, particularly in device-level identification of identical platforms. This paper proposes a unified RF fingerprinting framework that integrates statistical features, multi-resolution time–frequency analysis, and deep learning representations. A multi-segment high order cumulant method (HOCv2) is introduced to preserve temporal dynamics, while wavelet based RF-DNA features capture multi-scale characteristics. These handcrafted features are fused to enhance discriminative capability. In parallel, time–frequency spectrograms are employed to train a lightweight convolutional neural network (CNN) for automatic feature extraction. Experimental results on seven identical UAVs show that feature fusion achieves 91.04% accuracy, while the CNN attains 92.37% accuracy with a macro AUC of 0.9949 and high computational efficiency. The results demonstrate the effectiveness and complementarity of the proposed unified framework for robust UAV RF fingerprinting.

Key words - UAV Identification; RF Fingerprinting; High order Cumulants; Wavelet RF-DNA; Convolutional Neural Networks

1. Introduction

The rapid development of unmanned aerial vehicles (UAVs) in recent years has led to widespread applications across various domains, including civilian, commercial, and defense sectors. However, this growth also introduces significant challenges related to airspace monitoring, information security, and spectrum management. Accurate detection and identification of UAVs, particularly in complex environments or under conditions where visual observation is limited, have therefore become critical research problems [1], [2].

Traditional approaches based on computer vision or radar systems have been extensively studied; however, their performance often degrades under adverse conditions such as low illumination, unfavorable weather, or when dealing with small sized and low altitude UAVs [3]. In this context, radio frequency (RF) fingerprinting has emerged as an effective alternative by exploiting hardware imperfections and unique transmission characteristics of emitting devices, enabling the differentiation of UAVs even when operating under the same communication standards [4], [5].

However, RF signals emitted by UAVs exhibit complex characteristics, including non stationarity, time varying behavior due to motion dynamics (particularly in hovering states), and strong influences from wireless propagation channels. These factors degrade the discriminability of traditional features and pose significant challenges for machine learning models in terms of generalization [6]. Recent studies have proposed direct exploitation of in phase and quadrature (IQ) signals, typically using one dimensional convolutional neural networks (1D-CNNs), to automatically learn feature representations from raw data [7]. Although these approaches have demonstrated high performance in certain experimental settings, they often heavily depend on network architecture design and require large scale datasets, while lacking physical interpretability.

In parallel, handcrafted feature based approaches continue to play an important role in RF fingerprinting. High order cumulants (HOC) have been widely used to characterize the non Gaussianity and nonlinear properties of RF signals [8], [9], while RF-DNA (Radio Frequency Distinct Native Attribute) features based on wavelet transforms enable multi scale time–frequency analysis, thereby capturing localized and effective signal characteristics [10], [11]. However, these methods typically exploit only a single aspect of the signal and do not fully leverage the complementary nature among different feature representations.

From the above analysis, it can be observed that a research gap exists in developing a unified RF fingerprinting framework capable of effectively integrating statistical features, multi scale representations, and modern deep learning approaches. In particular, jointly exploiting temporal dynamics, nonlinearity, and structural properties of RF signals has the potential to significantly enhance UAV identification performance.

In this paper, we propose a unified RF fingerprinting framework for UAV identification, in which multiple complementary approaches are integrated within a single processing pipeline. Specifically, a segmented high order cumulant method (HOCv2) is developed to preserve temporal variations in RF signals, while wavelet based RF-DNA features are employed to capture multi scale characteristics. These two groups of handcrafted features are then fused to leverage their complementary strengths. In addition, a lightweight convolutional neural network (CNN) is designed to learn discriminative representations from time–frequency spectrograms obtained via short time Fourier transform (STFT).

Experimental results on a UAV RF dataset consisting of 13,893 samples from seven DJI Matrice 100 devices of the same model demonstrate that the proposed method

achieves high and stable performance. In particular, the CNN model attains an accuracy of up to 92.37%, outperforming traditional feature based approaches, while maintaining low computational complexity and enabling real time inference. These results highlight the practical potential of the proposed framework for spectrum surveillance and UAV identification in real world scenarios.

2. Dataset and Methods

2.1. UAV RF Dataset and Research Methodology

2.1.1. UAV RF Dataset

This study utilizes a UAV RF signal dataset publicly released by Soltani et al. in IEEE Transactions on Vehicular Technology [6]. The dataset was collected in an anechoic chamber, where seven DJI Matrice 100 UAVs of the same model acted as transmitters. On the receiver side, signals were captured using an Ettus USRP X310 equipped with a UBX-160 daughterboard, operating in the downlink band at 10 MHz, corresponding to the transmission frequency of the UAVs. The data were acquired at four different distances: 6ft, 9ft, 12ft, and 15ft. At each distance, signal recordings were performed for approximately 2 seconds, followed by a 10-second pause, and repeated three additional times, resulting in four non overlapping bursts per UAV at each distance. After removing interference from adjacent frequency bands, each burst yields approximately 140 non overlapping samples. In total, for seven UAVs across four distances with four bursts per distance, the final dataset contains approximately 13,000 samples, with an average length of about 92,000 complex I/Q samples per recording. A notable characteristic of this dataset is that all UAV signals were collected under hovering flight conditions.

In this study, the signals are processed in the form of complex baseband I/Q data stored according to the SigMF (Signal Metadata Format) standard [12]. According to the SigMF structure, each recording consists of a data section containing I/Q samples and a metadata section describing acquisition parameters such as sampling rate, center frequency, and other relevant system information. The use of the SigMF format standardizes RF data management and ensures reproducibility, while facilitating efficient processing in MATLAB.

For model training and evaluation, invalid samples are removed, and the remaining 13,893 samples are divided into training, validation, and testing sets with a ratio of 60%, 20%, and 20%, respectively. From a mathematical perspective, each RF signal recording in the baseband is represented as a complex sequence:

$$\boldsymbol{x}[n] = \boldsymbol{I}[n] + j\boldsymbol{Q}[n], n = 0, ..., N-1, \quad (1)$$

where $\boldsymbol{I}[n]$ and $\boldsymbol{Q}[n]$ denote the in-phase and quadrature components at time index n, and N is the total number of samples in the recording. This representation forms the foundation for subsequent processing steps in the study, including high order statistical feature extraction, wavelet based feature analysis, and time–frequency representation construction for deep learning models.

2.1.2. Research Methodology

Based on the aforementioned UAV RF dataset, this study develops a unified processing pipeline to simultaneously exploit three complementary groups of features, including statistical features, multi scale time–frequency representations, and time–frequency image representations for deep learning. The central idea of this work is not to rely on a single type of feature for UAV identification, but rather to construct a unified framework that enables both comparison and integration of multiple approaches on the same dataset. The overall processing pipeline consists of the following main steps:

- Step 1: Read and standardize I/Q data from the SigMF format. RF signals are first segmented and converted into complex sequences $\boldsymbol{x}[n]$. At this stage, acquisition parameters such as sampling rate and metadata information are synchronized with the signal data to support subsequent training and evaluation processes.

- Step 2: Signal segmentation and preprocessing. For recordings with large durations, the signals are segmented or divided into appropriate portions depending on the requirements of each feature extraction branch. For handcrafted feature extraction, the raw or normalized signals are directly fed into HOC, HOCv2, or wavelet based RF-DNA extraction modules. For the deep learning branch, the signals are transformed into time–frequency representations using short time Fourier transform (STFT) to generate spectrograms of fixed size.

- Step 3: Handcrafted feature extraction. Three main types of features are investigated, including conventional high order cumulants (HOC), the proposed segmented cumulant method (HOCv2), and wavelet based RF-DNA features.

- Step 4: Feature fusion. The HOCv2 and wavelet RF-DNA feature sets are concatenated to form an 80-dimensional fused feature vector. This design is motivated by the observation that HOCv2 effectively captures temporal nonlinear statistical variations, while wavelet RF-DNA encodes multi resolution energy distributions. Therefore, combining these feature groups enhances discriminative capability compared to using each feature set individually.

- Step 5: Dataset construction for deep learning. In parallel with the handcrafted feature branch, the I/Q signals are transformed into spectrograms using STFT. The resulting time–frequency images are then normalized to a fixed size and used as inputs to the proposed CNN model, enabling automatic learning of discriminative feature representations.

- Step 6: Model training and evaluation. For handcrafted and fused feature vectors, conventional classifiers, including support vector machines (SVM) with an RBF kernel, are employed and evaluated under the same 60%/20%/20% data splitting protocol to ensure fairness within the handcrafted feature branch. For the CNN-based branch, a separate 70%/15%/15% split is adopted to provide sufficient training samples for deep learning.

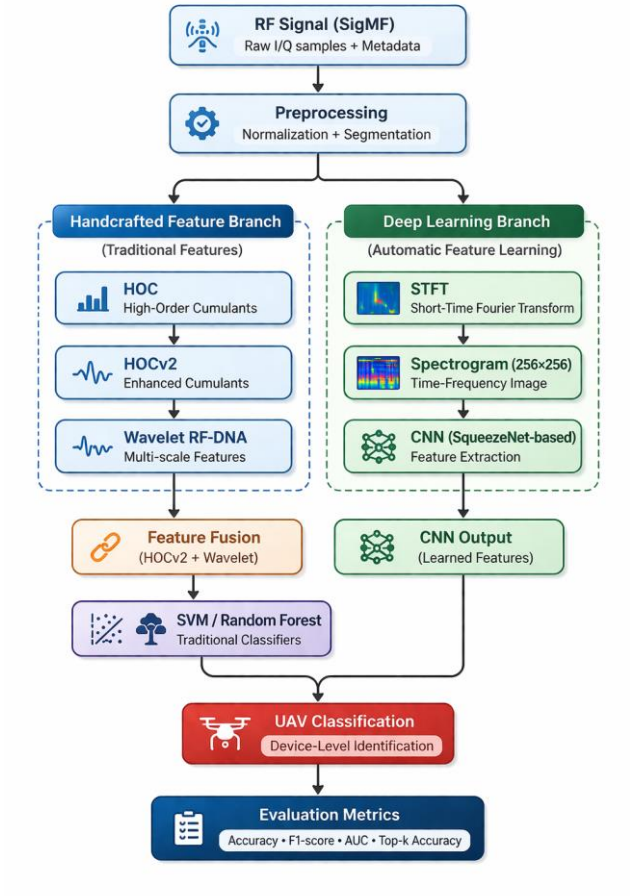


Figure 1. *Overall research pipeline of the proposed RF fingerprinting framework for UAV identification*

From a methodological perspective, the proposed pipeline is designed to address three key research questions. First, whether conventional statistical features such as HOC remain effective for UAV identification. Second, whether preserving temporal variations through HOCv2 and capturing multi scale characteristics via wavelet transforms can improve identification performance. Third, whether a deep learning model based on spectrogram representations can outperform handcrafted feature based methods while maintaining real time deployment capability. Organizing the workflow in a unified framework, as illustrated, ensures a systematic, transparent, and fair comparison among different approaches, thereby enhancing the reliability and practical relevance of the experimental results.

2.2. Signal Preprocessing and Time–Frequency Representation

2.2.1. Signal Normalization and Segmentation

After loading the signals from binary files and the SigMF dataset, the first preprocessing step is to convert the signals into complex vectors and perform amplitude normalization. Given the input signal $\boldsymbol{x}[n]$, a commonly used normalization is defined as:

$$\tilde{\boldsymbol{x}}[n] = \frac{\boldsymbol{x}[n]}{\sqrt{\frac{1}{N}\sum_{n=0}^{N-1}|\boldsymbol{x}[n]|^2 + \varepsilon}}, \quad (2)$$

where ε is a small constant introduced to avoid division by zero. This normalization reduces the impact of amplitude variations across different recordings and ensures that the extracted features primarily reflect intrinsic signal characteristics, rather than differences in received signal power. This normalization step is particularly important because the received signal power may vary due to factors such as transmission distance, UAV position fluctuations during hovering, and channel effects. Previous studies [6] have also indicated that hovering conditions can introduce significant temporal variations, leading to noticeable changes in the received signal distribution across different snapshots. RF recordings in the dataset are relatively long; therefore, directly processing the entire sequence would increase computational complexity while potentially overlooking local temporal variations of RF fingerprints. To address this issue, the signals are segmented into shorter segments. Let the segment length be L, the number of segments be S, and the stride be R.

The s segment is defined as:

$$x_s[l] = \tilde{x}[sR + l]; l = 0, ..., L-1; s = 0, ..., S-1, \quad (3)$$

with the condition $sR+L-1<N$. This segmentation step serves three main purposes. First, it reduces the computational complexity of feature extraction methods. Second, it enables the capture of short term temporal variations, which are particularly important in UAV hovering scenarios. Third, it increases the effective dataset size, which is beneficial for training machine learning models.

In this study, segmentation is used not only for constructing time–frequency representations but also as a key component of the proposed HOCv2 method. Specifically, instead of computing high order cumulants over the entire burst, HOCv2 computes them separately on multiple temporal segments and concatenates the resulting segment-level descriptors. This design helps preserve the temporal evolution and dynamic statistical structure of RF fingerprints.

2.2.2. STFT Transformation and Spectrogram Construction

Since UAV RF signals are inherently non stationary, conventional Fourier transform analysis is insufficient to capture temporal variations in frequency components. Therefore, this study employs the short time Fourier transform (STFT) to construct time–frequency representations. In non stationary signal analysis, STFT is a standard tool for evaluating the temporal evolution of spectral content. The STFT of a signal is defined as:

$$x(m,k) = \sum_{n=-\infty}^{\infty} \tilde{x}[n]\, w[n - mR]\, e^{-j2\pi kn/K}, \quad (4)$$

where $w[\cdot]$ is the analysis window, m denotes the frame index, R is the hop size, k is the discrete frequency index, and K is the number of FFT points.

From the STFT, the time–frequency energy representation, or spectrogram, is defined as:

$$S(m,k) = |X(m,k)|^2, \quad (5)$$

and in practice, a log scale representation is often used to compress the dynamic range:

$$S_{\log}(m,k) = 10\log_{10}\left(\left|X(m,k)\right|^2 + \delta\right), \quad (6)$$

where δ is a small constant introduced to avoid taking the logarithm of zero. The log-spectrogram representation is particularly useful when the signal spectrum exhibits a large dynamic range, as it enhances prominent spectral structures that carry discriminative information.

In UAV signal analysis, the spectrogram not only reflects the energy distribution of the transmitted signal but may also capture structures related to micro Doppler effects, which arise from mechanical vibrations of UAV components. These structures have been widely recognized as valuable sources of information for UAV detection and identification. After computing the spectrogram, the resulting time–frequency matrix is normalized and resized to a fixed dimension to serve as input to the CNN model. Specifically, the time–frequency representations are converted into images of size 256×256, ensuring consistency across samples for deep learning training.

2.3. High order Cumulant Features for RF Fingerprinting

2.3.1. Mathematical Foundation and HOC Feature Representation

In practical RF systems, signals transmitted from different devices often exhibit inherent variations due to hardware imperfections. These include local oscillator instabilities, I/Q imbalance, nonlinear distortions in power amplifiers, frequency offsets, and other hardware related effects, all of which manifest as statistical variations in the received signal. In this context, high order cumulants (HOCs) serve as powerful mathematical tools for characterizing the non Gaussianity and nonlinear properties of signals. Consequently, they have been widely employed in applications such as modulation classification and RF fingerprinting.

Considering the complex baseband signal $\boldsymbol{x}[n]$, higher order moments can be generally defined as expectations of powers of $\boldsymbol{x}[n]$ and its complex conjugate $\boldsymbol{x}^*[n]$. From these moments, cumulants are constructed to eliminate redundant contributions from lower order correlations. In this study, we focus on second and fourth order cumulants, as they are among the most widely used features for modulation classification and device identification due to their effectiveness in capturing nonlinear characteristics.

For normalized and zero mean signals, the basic moments are defined as:

$$\begin{aligned} &m_{20} = E\left\{\boldsymbol{x}^2[n]\right\}, m_{21} = E\left\{\left|\boldsymbol{x}[n]\right|^2\right\}, \\ &m_{40} = E\left\{\boldsymbol{x}^4[n]\right\}, m_{41} = E\left\{\boldsymbol{x}^3[n]\boldsymbol{x}^*[n]\right\}, \\ &m_{42} = E\left\{\left|\boldsymbol{x}[n]\right|^4\right\}. \end{aligned} \quad (7)$$

From the above moments, the corresponding cumulants are defined as:

$$\begin{aligned} &C_{20} = m_{20}, C_{21} = m_{21}, \\ &C_{40} = m_{40} - 3m_{20}^2, \\ &C_{41} = m_{41} - 3m_{20}m_{21}, \\ &C_{42} = m_{42} - \left|m_{20}\right|^2 - 2m_{21}^2. \end{aligned} \quad (8)$$

Here, C_{20} and C_{21} represent second order statistics of the signal, while C_{40}, C_{41}, C_{42} capture fourth order characteristics, which are sensitive to non Gaussianity and nonlinear effects. In this study, normalization with respect to m_{21} is applied to obtain dimensionless features:

$$\begin{aligned} &\tilde{C}_{20} = \frac{C_{20}}{m_{21}}, \tilde{C}_{21} = \frac{C_{21}}{m_{21}}, \\ &\tilde{C}_{40} = \frac{C_{40}}{m_{21}^2}, \tilde{C}_{41} = \frac{C_{41}}{m_{21}^2}, \\ &\tilde{C}_{42} = \frac{C_{42}}{m_{21}^2}. \end{aligned} \quad (9)$$

Since the signal is complex valued, each complex cumulant can be decomposed into its real and imaginary parts, thereby forming feature components. Accordingly, a basic HOC feature vector can be expressed as:

$$\mathbf{f}_{\text{HOC}} = \left[\Re\left(\tilde{C}_{20}\right), \Im\left(\tilde{C}_{20}\right), ..., \Re\left(\tilde{C}_{42}\right), \Im\left(\tilde{C}_{42}\right)\right]. \quad (10)$$

This vector captures important statistical distortions in the transmitted signal and forms the conventional HOC representation for each RF signal sample. One key advantage of HOC is its low computational complexity, independence from signal demodulation, and clear physical interpretability. Compared to image based deep learning methods, HOC is easier to interpret as it is directly associated with the nonlinear characteristics of the transmitter. However, this statistical nature also introduces a limitation: when cumulants are computed over an entire long burst, temporal variations in the signal may be averaged out or suppressed. As a result, the discriminative capability can be reduced, especially when distinguishing UAVs with similar RF fingerprints.

2.3.2. Proposed HOCv2 Method

In the UAV dataset considered in this study, the UAVs operate in hovering conditions, which introduce small positional fluctuations and cause the wireless channel to vary over time. These variations can lead to changes in the statistical structure of RF signals within short segments of each burst. When HOC features are computed over the entire burst, such local variations may be averaged out, thereby reducing the discriminative capability between UAVs. To address this limitation, this study proposes an extended HOC approach, referred to as Multi Segment Higher order Cumulants (HOCv2).

The main idea of the proposed method is to divide each RF signal into multiple shorter temporal segments and compute high order cumulants separately for each segment. This segmentation strategy better preserves the temporal dynamics of RF signals. Specifically, let an RF signal $\boldsymbol{x}[n]$

of length N be divided into S equal-length segments:

$$\boldsymbol{x}[n] = \left\{x_1[n], x_2[n], ..., x_S[n]\right\}, \quad (11)$$

where $x_s[n]$ denotes the s-th segment, s=1,2,…,S. For each segment $x_s[n]$, high order cumulants are computed using the formulations introduced previously. This results in a set of HOC features for each segment:

$$\mathbf{f}_s = \left\{C_{20}^{(s)}, C_{21}^{(s)}, C_{40}^{(s)}, C_{41}^{(s)}, C_{42}^{(s)}\right\}, \quad (12)$$

where $C_{pq}^{(s)}$ denotes the cumulant of order p,q computed on the s-th segment. After separating the real and imaginary parts of the complex cumulants, the feature vector for each segment is constructed as:

$$\mathbf{f}_{\text{HOC}} = \left[\Re\left(C_{20}^{(s)}\right), \Im\left(C_{20}^{(s)}\right), ..., \Re\left(C_{42}^{(s)}\right), \Im\left(C_{42}^{(s)}\right)\right] \quad (13)$$

Finally, the feature vectors from all segments are concatenated to form the overall HOCv2 feature vector:

$$\mathbf{f}_{\text{HOCv2}} = \left\{\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_S\right\}. \quad (14)$$

If each segment produces d HOC features, the final feature vector has a dimensionality of:

$$D = S \times d. \quad (15)$$

In this study, each RF burst is divided into multiple temporal segments, and high order cumulants are computed for each segment. This approach enables the capture of time-varying statistical characteristics of RF signals, thereby enhancing the discriminative capability between similar UAVs. Compared to conventional HOC, the proposed HOCv2 method offers several key advantages: it preserves the dynamic structure of RF signals, increases feature resolution, and improves the ability to distinguish between RF devices.

Experimental results demonstrate that HOCv2 significantly improves UAV classification performance compared to traditional HOC, confirming that exploiting the temporal structure of RF fingerprints is an effective strategy for UAV identification systems.

2.4. Wavelet based RF-DNA Method

In addition to high order statistical features, multi resolution time–frequency analysis methods, particularly wavelet transforms, have been proven effective for RF fingerprint feature extraction. Among these, wavelet based RF-DNA is a widely adopted approach for capturing localized and multi scale characteristics of RF signals.

Given the complex baseband RF signal $\boldsymbol{x}[n]$, the discrete wavelet transform (DWT) is defined as:

$$W_{j,k} = \sum_n \boldsymbol{x}[n] \psi_{j,k}(n), \quad (16)$$

where, $\psi_{j,k}(n)$ denotes the mother wavelet function, j is the decomposition level, and k represents the time index. In practice, the DWT is implemented through multi level decomposition, where the signal is separated into approximation coefficients (A), representing low frequency components, and detail coefficients (D), representing high-frequency components. After L levels of decomposition, the signal is represented by the set of coefficients:

$$\left\{A_L, D_L, D_{L-1}, ..., D_1\right\}. \quad (17)$$

These wavelet coefficients capture spectral characteristics of the RF signal across multiple frequency scales. In RF fingerprinting, they reflect hardware induced impairments such as nonlinear distortions and frequency offsets, which manifest as distinctive patterns in the signal spectrum. Following wavelet decomposition, RF-DNA features are constructed by computing statistical descriptors of the wavelet coefficients.

In this study, four primary statistical features are extracted for each wavelet band.

- Energy:

$$E = \sum_n \left|c[n]^2\right|, \quad (18)$$

- Mean:

$$\mu = \frac{1}{N}\sum_n c[n], \quad (19)$$

- Standard deviation:

$$\sigma = \sqrt{\frac{1}{N}\sum_n \left(c[n] - \mu\right)^2}, \quad (20)$$

- Kurtosis:

$$k = \frac{E\left[\left(c[n] - \mu\right)^4\right]}{\sigma^4}, \quad (21)$$

where $c[n]$ denotes the wavelet coefficients corresponding to a specific frequency band. These statistical measures are computed across multiple wavelet bands and concatenated to form the RF-DNA feature vector:

$$\mathbf{f}_{\text{RF-DNA}} = \left[A_E, A_\mu, A_\sigma, A_\kappa, D_1^E, D_1^\mu, D_1^\sigma, D_1^\kappa, ...\right] \quad (22)$$

This feature vector characterizes the spectral properties of RF signals across multiple frequency scales and can be used to distinguish between different RF devices. Compared to HOC features, RF-DNA offers several notable advantages:

- It captures multi scale spectral structures of RF signals, enabling the detection of hardware induced distortions across different frequency bands.

- It provides improved representation of non stationary signals due to the multi resolution nature of the wavelet transform.

- It complements statistical features, as RF-DNA focuses on spectral characteristics, whereas HOC describes statistical structures of the signal.

Therefore, in this study, RF-DNA is utilized as a complementary feature source to HOCv2. The combination of these features enables the construction of a richer RF fingerprint representation, thereby improving UAV classification performance.

2.5. Feature Fusion of HOCv2 and Wavelet RF-DNA

RF fingerprinting features can be extracted from multiple domains of RF signals, including statistical, temporal, frequency, and time–frequency domains. Each feature type

captures different aspects of the RF signal. In particular, HOCv2 characterizes the statistical structure of the signal, while wavelet RF-DNA features describe its spectral properties. These two feature groups are inherently complementary. While HOCv2 focuses on statistical variations caused by nonlinear distortions in RF transmitters, RF-DNA emphasizes spectral structures across multiple frequency scales. Therefore, combining these two feature types can provide a richer RF fingerprint representation and improve UAV classification performance.

In this study, a unified fusion approach is proposed, where feature vectors extracted from HOCv2 and RF-DNA are directly concatenated to form a single comprehensive feature vector. Assume that:

$$\mathbf{f}_{\text{HOCv2}} = \left[h_1, h_2, \ldots, h_{d_1} \right] \in \mathbb{R}^{d_1}, \tag{23}$$

where d_1 represents the dimensionality of the HOCv2 descriptor, and each element h_i corresponds to a higher order cumulant based RF statistical feature extracted from the UAV signal. Similarly, the Wavelet RF-DNA descriptor is represented as:

$$\mathbf{f}_{\text{RF-DNA}} = \left[w_1, w_2, \ldots, w_{d_2} \right] \in \mathbb{R}^{d_2}, \tag{24}$$

where d_2 denotes the dimensionality of the Wavelet RF-DNA descriptor, and each element w_i corresponds to a wavelet domain statistical RF fingerprint feature obtained from multi resolution time–frequency analysis. The fused feature vector is then constructed by concatenating the two descriptors:

$$\mathbf{f}_{\text{fusion}} = \left[\mathbf{f}_{\text{HOCv2}}^T, \mathbf{f}_{\text{RF-DNA}}^T \right]^T \in \mathbb{R}^{d_1+d_2}, \tag{25}$$

with total dimensionality $d=d_1+d_2$. In this study, the HOCv2 feature vector has approximately 60 dimensions, while the wavelet RF-DNA feature vector has around 20 dimensions, resulting in a fused feature vector of approximately $d\approx80$.

Before training classification models, the feature vectors are normalized to ensure consistent scaling and to prevent any feature group from dominating the learning process. In this study, Z-score normalization is applied:

$$\hat{x} = \frac{x-\mu}{\sigma}, \tag{26}$$

where μ and σ denote the mean and standard deviation of each feature computed from the training set. After normalization, the fused feature vectors are used as inputs to conventional machine learning classifiers, including Support Vector Machine (SVM), for UAV classification.

2.6. Lightweight CNN-Based RF Fingerprinting Framework

In addition to handcrafted features, this study develops a deep learning model to automatically learn discriminative representations from time–frequency images. The use of CNNs on spectrograms is well-suited to the non stationary nature of RF signals, as STFT highlights localized structures in both time and frequency domains. This approach has been widely explored in signal and UAV identification tasks [13].

The input to the proposed CNN model is a spectrogram image of size 256×256×3, constructed from the I/Q signals after preprocessing and time–frequency transformation. For an input sample $\boldsymbol{X}\in \mathbb{R}^{256\times256\times3}$, the network learns a nonlinear mapping that transforms the input spectrogram into a posterior probability vector over seven UAV classes through the softmax function:

$$\boldsymbol{y} = \text{softmax}\left(z_\theta\left(\boldsymbol{X} \right) \right), \boldsymbol{y} \in \left[0,1 \right]^7, \sum_{c=1}^{7} y_c = 1 \tag{27}$$

where, θ denotes the set of trainable network parameters, $z_\theta(\boldsymbol{X})$ represents the output logit vector before the softmax layer, and $\boldsymbol{y}$ is the posterior probability vector corresponding to the seven UAV classes. The predicted UAV label is determined using the maximum posterior probability criterion.

2.6.1. Proposed CNN Architecture

The proposed model is a lightweight CNN based on the SqueezeNet architecture, as illustrated in Figure 2, consisting of 92 layers and 99 connections. Fire modules are adopted as the core building blocks for feature extraction [14]. The SqueezeNet architecture is designed to achieve high accuracy with a compact model size, making it suitable for resource constrained systems.

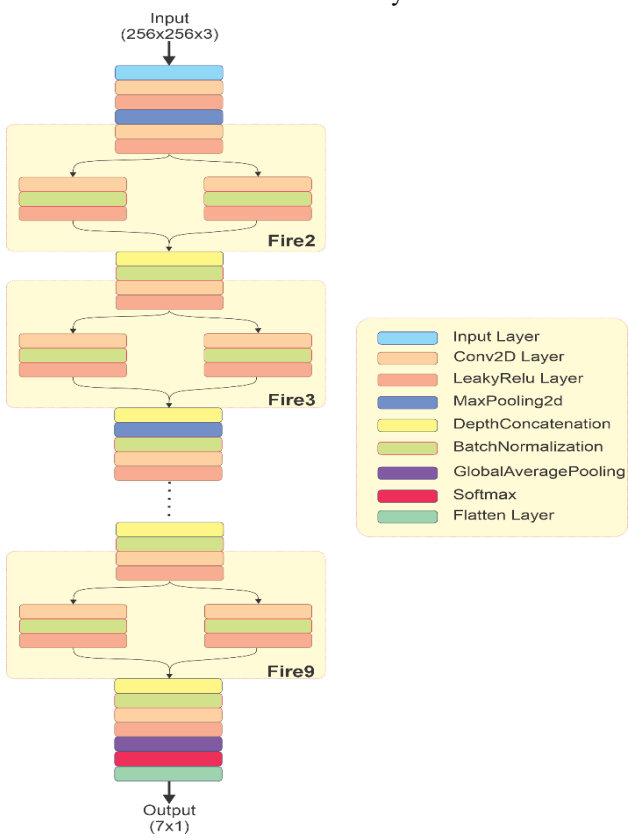


Figure 2. *Architecture of the proposed CNN model*

At the input stage, the spectrogram image is first processed by a convolutional layer with a 3×3 kernel, 64 filters, and a stride of 2. This layer reduces the spatial dimensions while enhancing feature representation. A LeakyReLU activation function with a negative slope of 0.01 is applied, followed by a 2×2 max-pooling layer (stride 2) to further reduce resolution and improve

robustness to local variations. The core of the network consists of a sequence of Fire modules from Fire2 to Fire9.

Each Fire module is composed of two stages: a squeeze layer and an expand layer. In the squeeze stage, a 1×1 convolution is used to reduce the number of feature channels, with channel sizes of 16, 16, 32, 32, 48, 48, 64, and 64 corresponding to Fire2 through Fire9, respectively. The output of the squeeze layer is then fed into two parallel branches in the expand stage: one branch uses 1×1 convolutions, while the other employs 3×3 convolutions with padding to preserve spatial dimensions. The number of filters in the expand layers increases progressively across the network, with values of 64, 64, 128, 128, 192, 192, 256, and 256 for each Fire module. The outputs of the two branches are concatenated along the channel dimension (depth concatenation), enhancing the network's ability to capture multi scale and multi level features.

After each convolutional layer, batch normalization is applied, followed by a LeakyReLU activation function, to stabilize feature distributions, accelerate convergence, and enhance nonlinear representation capability. Max-pooling layers with a 2×2 kernel are inserted after Fire3 and Fire5 to reduce spatial dimensions and enlarge the receptive field of the network. At the final stage, the output of Fire9 is passed through a 1×1 convolutional layer with 7 output channels, corresponding to the number of target classes. Subsequently, a global average pooling layer is applied to convert the feature maps into a global feature vector, thereby reducing the number of parameters compared to traditional fully connected layers. Finally, a softmax layer is used to estimate the class probabilities for classification.

Overall, the proposed architecture effectively leverages the combination of 1×1 and 3×3 convolutions within Fire modules to balance computational efficiency and feature learning capacity. The model achieves a favorable trade-off between performance and complexity, making it particularly suitable for classification tasks involving time–frequency spectrograms and RF signal based device identification.

2.6.2. Loss Function and Training Strategy

Since the problem is formulated as a single-label classification task with 7 classes, the cross-entropy loss function is employed on the softmax outputs. Given a mini-batch of size $\mathbf{B}$, let the softmax outputs be $\mathbf{Y} \in \mathbb{R}^{7\times\mathbf{B}}$ and the corresponding ground-truth labels be $\mathbf{T} \in \mathbb{R}^{7\times\mathbf{B}}$. The loss function is defined as:

$$\mathcal{L} = -\frac{1}{B}\sum_{i=1}^{B}\sum_{c=1}^{7} T_{c,i} \log\ Y_{c,i} + \varepsilon \ , \qquad (28)$$

where ε is a small constant to ensure numerical stability. Cross entropy is well suited for multi class classification problems, as it directly measures the discrepancy between predicted probabilities and ground-truth labels. In this study, the model is encouraged to assign higher probabilities to the correct UAV class while suppressing incorrect predictions.

The model is trained using a mini-batch gradient descent scheme on a GPU. The dataset is divided into training, validation, and testing sets, and the training samples are shuffled at each epoch to mitigate overfitting. The Adam optimizer is employed with an initial learning rate of 10^{-5}, first order moment coefficient β_1=0.9, and second order moment coefficient β_2=0.999. Adam is an adaptive optimization method that combines the advantages of momentum and adaptive learning rates, enabling efficient convergence on noisy and non stationary data. A mini-batch size of 16 is used. At each iteration, gradients of the loss function with respect to network parameters are computed via backpropagation and used to update the parameters using Adam. After each epoch, the model is evaluated on the validation set, and the best-performing model is saved as a checkpoint. This training strategy helps reduce overfitting and ensures that the learned model generalizes well to unseen data, rather than overfitting to the training set.

3. Experimental Results and Evaluation

3.1. Experimental Setup and Evaluation Protocol

In this study, experiments are conducted on an RF signal dataset collected from seven DJI Matrice 100 UAV devices of the same model, where each device is treated as a separate class. Therefore, the problem is formulated as a device level identification task based on RF fingerprints, rather than classification by UAV type. The input signals consist of wideband I/Q data conforming to the SigMF standard. After preprocessing, the signals are segmented into shorter bursts of fixed length. These bursts serve as input samples for the entire processing pipeline, including both handcrafted feature extraction methods and deep learning models. For machine learning models based on HOC, HOCv2, wavelet RF-DNA, and feature fusion, the dataset is split into training, validation, and testing sets using a 60/20/20 ratio. From a total of 13,893 samples, the training set contains 8,337 samples, while the validation and testing sets each contain 2,778 samples.

The training results show that the HOC, HOCv2, wavelet, and feature-fusion methods were evaluated under the same data splitting protocol to ensure fairness and consistency in performance comparison. For the deep learning branch based on RF fingerprint images, the CNN model was trained using input images of size 256 × 256 × 3, which were generated from RF signals after time–frequency transformation and normalization. The dataset was divided according to a 70%/15%/15% ratio, resulting in 9,725 training samples, 2,084 validation samples, and 2,084 testing samples. The training set was used to update the model parameters, the validation set was used to monitor the training process and select the best performing model, while the testing set was used only for the final evaluation. The output layer of the model consisted of 7 classes, corresponding to the seven UAV categories considered in this study. This setting ensures that both the handcrafted feature based methods and the CNN model were evaluated under the same multi-class classification task. To evaluate model performance, multiple metrics are employed. First, the fundamental metric is classification

accuracy, defined as:

$$Accuracy = \frac{N_{correct}}{N_{total}}, \tag{29}$$

where N_{correct} is the number of correctly predicted samples and N_{total} is the total number of samples in the test set. This metric provides a primary basis for comparing the performance of different methods. In addition to accuracy, precision, recall, and F1-score are used for each UAV class. For a given class, these metrics are defined as:

$$\begin{aligned} Precision &= \frac{TP}{TP+FP}, \\ Recall &= \frac{TP}{TP+FN}, \\ F1 &= \frac{2\times Precision\times Recall}{Precision+Recall}. \end{aligned} \tag{30}$$

Here, TP, FP, and FN denote the numbers of true positives, false positives, and false negatives, respectively. Precision reflects the reliability of positive predictions, while recall measures the ability to correctly identify samples belonging to a given class. The F1-score is the harmonic mean of precision and recall. For multi class classification tasks, these metrics are aggregated using macro-averaging and weighted averaging to account for class imbalance and to provide a comprehensive evaluation of model performance.

For the CNN model, additional advanced metrics are considered. Specifically, Top-k accuracy is used to measure whether the correct label appears within the top-k predicted classes with the highest probabilities. In addition, Cohen's kappa coefficient is employed to quantify the agreement between predicted labels and ground-truth labels beyond chance. Given a confusion matrix $\mathbf{C}$, the kappa coefficient is defined as:

$$\kappa = \frac{p_0 - p_e}{1 - p_e}, \tag{31}$$

where p_0 is the observed accuracy and p_e is the expected accuracy due to random chance. A value of κ closer to 1 indicates a higher level of agreement.

To evaluate the quality of the predicted probability distributions, AUC and Brier score are employed. AUC measures the model's ability to discriminate between one class and the others, while the Brier score quantifies the mean squared error between predicted probabilities and ground-truth labels. A lower Brier score indicates better-calibrated predictions. In addition to accuracy based metrics, the study also considers model complexity and deployment feasibility. The experimental results show that the proposed CNN model has approximately 0.736 million parameters and achieves an inference speed of about 142 FPS, demonstrating strong potential for real time UAV identification applications. All experiments were implemented using MATLAB R2024a and executed on a workstation equipped with an Intel Core i7 processor, 32 GB RAM, and an NVIDIA GeForce RTX 3060 GPU with 12 GB of dedicated memory. The proposed CNN model was trained and evaluated under GPU acceleration to ensure efficient model optimization and inference.

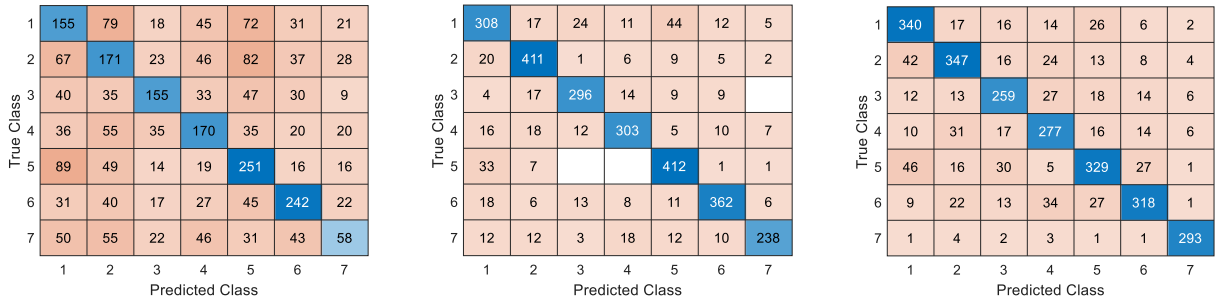
3.2. Results of Handcrafted Feature based Methods

To evaluate the effectiveness of handcrafted features in UAV identification based on RF fingerprinting, this study conducts a comparative analysis of three feature extraction methods: conventional HOC, the proposed HOCv2, and wavelet based RF-DNA, under the same training and testing protocol. Quantitative results for each UAV class are presented in Table 1, while the corresponding confusion matrices are illustrated in Figure 3. Evaluating all three methods under a unified experimental setup allows for a clear comparison of their respective contributions and effectiveness in distinguishing UAV devices of the same model.

Overall, the results indicate that conventional HOC achieves relatively low performance, with an overall test accuracy of 43.27%, highlighting the limitations of statistical representations computed over entire signal bursts. As shown in Table 1, the precision, recall, and F1-score of this method are consistently low and exhibit significant variation across different UAV classes. Specifically, the F1-scores are 0.349 for UAV-1, 0.365 for UAV-2, 0.490 for UAV-3, 0.449 for UAV-4, 0.494 for UAV-5, and 0.574 for UAV-6, before dropping sharply to 0.242 for UAV-7. This substantial degradation suggests that HOC features, when computed over the entire signal, fail to preserve local temporal variations, which are critical in UAV hovering scenarios. In other words, the accumulation of statistical information over long bursts tends to average out subtle hardware induced differences in RF fingerprints, thereby reducing the discriminative capability between UAVs with similar hardware characteristics. This observation is further confirmed by the confusion matrix shown in Figure 3(a), where significant off-diagonal elements are present, indicating a high degree of inter-class confusion and limited class separability.

Table 1. *Comparison of Precision, Recall, and F1-score across UAV classes for feature based methods*

Class	HOC			HOCv2			Wavelet RF-DNA		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
UAV-1	0.331	0.368	0.349	0.749	0.732	0.740	0.739	0.808	0.772
UAV-2	0.353	0.377	0.365	0.842	0.905	0.873	0.771	0.764	0.768
UAV-3	0.546	0.444	0.490	0.848	0.848	0.848	0.734	0.742	0.738
UAV-4	0.440	0.458	0.449	0.842	0.817	0.829	0.721	0.747	0.734
UAV-5	0.446	0.553	0.494	0.821	0.907	0.862	0.765	0.725	0.744
UAV-6	0.578	0.571	0.574	0.885	0.854	0.869	0.820	0.750	0.783
UAV-7	0.333	0.190	0.242	0.919	0.780	0.844	0.936	0.961	0.948



a) Conventional HOC *b) Proposed HOCv2* *c) Wavelet RF-DNA*

Figure 3. *Confusion matrices on the test set for conventional feature based methods*

Compared to conventional HOC, HOCv2 demonstrates a significant improvement in classification performance, achieving an overall test accuracy of 83.87%. As shown in Table 1, the performance gains are consistent across most UAV classes. Specifically, the F1-score for UAV-1 increases from 0.349 to 0.740, UAV-2 from 0.365 to 0.873, UAV-3 from 0.490 to 0.848, UAV-4 from 0.449 to 0.829, UAV-5 from 0.494 to 0.862, and UAV-6 from 0.574 to 0.869, while UAV-7 exhibits a substantial improvement from 0.242 to 0.844. These improvements validate the core hypothesis behind HOCv2: by segmenting the signal and computing cumulants on each segment rather than over the entire burst, temporal statistical variations of RF signals are effectively preserved.

As a result, the resulting feature vectors not only capture non Gaussianity and nonlinear characteristics of the transmitter but also encode the dynamic structure of RF fingerprints over time. The confusion matrix in Figure 3(b) further supports this observation, showing a more concentrated distribution along the main diagonal compared to conventional HOC. This indicates that HOCv2 significantly reduces inter-class confusion and enhances class separability among UAV devices.

For the wavelet based RF-DNA method, the model achieves an overall accuracy of 77.86%, which is significantly higher than conventional HOC but still lower than HOCv2. This result indicates that multi resolution time–frequency representations are effective for RF fingerprinting tasks. Wavelet features capture energy distributions across multiple frequency scales, thereby highlighting hardware induced distortions and localized spectral variations. As reported in Table 1, the F1-scores of this method are 0.772 for UAV-1, 0.768 for UAV-2, 0.738 for UAV-3, 0.734 for UAV-4, 0.744 for UAV-5, 0.783 for UAV-6, and a notably high 0.948 for UAV-7. These results suggest that RF-DNA is particularly effective in modeling certain UAV classes. However, the performance remains inconsistent across classes. Several UAVs exhibit significantly lower F1-scores compared to HOCv2, particularly UAV-2, UAV-3, UAV-4, UAV-5, and UAV-6. The confusion matrix in Figure 3(c) shows that, although wavelet features improve class separability compared to conventional HOC, substantial off-diagonal confusion still persists, indicating that the model has not fully resolved inter-class ambiguity as effectively as HOCv2.

From Table 1, an important trend can be observed: HOCv2 outperforms wavelet RF-DNA on most UAV classes, except for UAV-7, where the wavelet based method achieves superior performance. This observation suggests that the two feature groups capture different aspects of RF fingerprints. Specifically, HOCv2 is more effective in modeling dynamic statistical structures and time-varying nonlinear distortions, while wavelet RF-DNA excels at representing multiscale spectral energy distributions. This distinction indicates that the two feature types are not purely competitive but inherently complementary.

From a signal modeling perspective, the results in this section also reveal an important insight: in the task of identifying UAVs of the same model, the hardware induced differences between devices are relatively subtle. Therefore, feature representations must be sufficiently sensitive to capture fine-grained variations, while also being robust enough to avoid being dominated by channel effects. Conventional HOC fails to meet this requirement due to its reliance on global statistical representations. In contrast, RF-DNA achieves notable improvements by exploiting multi resolution information; however, it remains less effective in capturing short term temporal variations caused by hovering induced dynamics. As a result, among the handcrafted feature approaches, HOCv2 emerges as the most effective method. It also serves as a crucial foundation for constructing the fused feature representation presented in the next section.

3.3. Results of the combined feature extraction methods HOCv2 and Wavelet

Based on the analyses in the previous section, it can be observed that HOCv2 and Wavelet RF-DNA exploit two complementary aspects of RF signals. Specifically, HOCv2 captures higher order statistical structures and temporal variations, whereas RF-DNA focuses on spectral energy representations with multi resolution analysis in the time–frequency domain. From this perspective, this study proposes a feature fusion strategy to simultaneously leverage the advantages of both feature sets. The feature vectors, after normalization, are input into an SVM classifier with an RBF kernel. The experimental results are presented in Table 2 and Figure 4.

The results show that the fusion based method achieves a test accuracy of 91.04%, which is the highest among the handcrafted feature based approaches. Compared to

HOCv2 (83.87%) and Wavelet RF-DNA (77.86%), the improvements are approximately 7.17% and 13.18%, respectively. This is a significant outcome, demonstrating that combining features from different representations not only provides intuitive benefits but also delivers substantial quantitative improvements in UAV classification tasks.

A detailed class-wise analysis presented in Table 2 indicates that the fusion method achieves high and stable performance across all seven UAV classes. Specifically, the F1-scores for the classes are 0.871 (UAV-1), 0.905 (UAV-2), 0.945 (UAV-3), 0.890 (UAV-4), 0.908 (UAV-5), 0.922 (UAV-6), and 0.948 (UAV-7). Notably, all classes attain F1-scores greater than 0.87, indicating consistent classification performance across all UAV types, without the severe performance degradation observed in more challenging classes in the traditional HOC approach. In particular, classes that previously exhibited lower performance, such as UAV-1 and UAV-7 (as discussed in Section 3.2), show considerable improvement, reflecting enhanced discriminative capability in the fused feature space.

Table 2. *UAV classification performance of the HOCv2–Wavelet fusion method*

Class	Precision	Recall	F1-score
UAV-1	0.875	0.867	0.871
UAV-2	0.922	0.888	0.905
UAV-3	0.953	0.937	0.945
UAV-4	0.856	0.927	0.890
UAV-5	0.877	0.941	0.908
UAV-6	0.954	0.891	0.922
UAV-7	0.963	0.934	0.948

The confusion matrix in Figure 4 provides further visual evidence of this improvement. Compared to the confusion matrix in Figure 3, the diagonal elements in Figure 4 become more pronounced and dominant, while the off-diagonal elements are significantly reduced. This indicates that UAV samples are more tightly clustered in the fused feature space, and the decision boundaries between classes are more clearly defined. Moreover, the level of confusion between UAV classes with similar RF fingerprints is substantially decreased, demonstrating that the feature fusion approach enhances discriminative capability, particularly in challenging classification scenarios. From a feature representation perspective, these results can be explained as follows. HOCv2 provides information on non Gaussian and nonlinear statistical deviations of RF signals during propagation, making it sensitive to channel variations and UAV dynamics. In contrast, Wavelet RF-DNA describes the energy distribution of the signal at multiple scales, thereby capturing features related to hardware structure and radiation characteristics of the device. When these two sources of information are integrated, the resulting fused feature vector becomes not only more informative but also complementary, enabling the classifier to exploit both statistical and time–frequency domain information. This leads to improved discriminative capability among UAVs with subtle hardware level differences, particularly in scenarios where the signal is affected by channel variations.

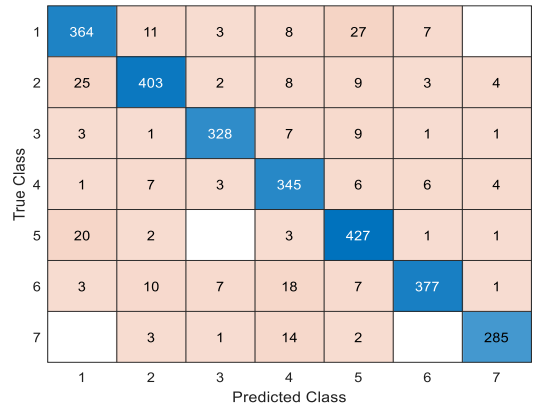


Figure 4. *Confusion matrix of the HOCv2–Wavelet combined method*

Another important aspect is the use of SVM with an RBF kernel in the fusion based feature space. The RBF kernel enables mapping data into a higher-dimensional nonlinear feature space, thereby constructing more flexible decision boundaries between classes. With feature vectors enriched by both HOCv2 and RF-DNA information, the SVM-RBF model can effectively exploit the nonlinear structure of the data, leading to improved classification performance. This also explains why the fusion method, despite not relying on deep learning, still achieves high accuracy and approaches the performance of CNN based models presented in the following section. Compared with the results in Section 3.2, it can be observed that the fusion method not only improves overall accuracy but also enhances the balance of performance across different classes. While HOCv2 tends to perform better for classes sensitive to temporal variations, and RF-DNA excels in classes with distinct spectral characteristics, the fusion approach successfully combines these advantages to achieve consistent performance across the entire dataset. This is particularly important in practical UAV identification systems, where robustness is required not only in terms of high accuracy but also in maintaining stable performance across all target classes.

However, despite achieving high performance, the fusion based method still depends on manual feature engineering, including the selection of parameters such as HOCv2 order, wavelet type, and resolution levels. These factors may affect the generalization capability when applied to different datasets or scenarios. Therefore, a natural direction is to transition toward deep learning models that can automatically learn feature representations directly from data, thereby reducing reliance on handcrafted features. This serves as the motivation for further investigating CNN based models on spectrogram representations in the following section.

3.4. Results of the proposed CNN based methods

In addition to handcrafted feature based approaches, this study further investigates a deep learning based approach, where a CNN model is trained directly on STFT spectrogram representations of RF signals. Unlike handcrafted features that capture only specific aspects of the signal, CNN models are capable of learning rich feature

representations in the time–frequency domain, including structural patterns, energy variations over time, and micro Doppler signatures. The experimental results are presented in Figures 5 and 6, along with quantitative evaluations in Tables 3, 4, and 5.

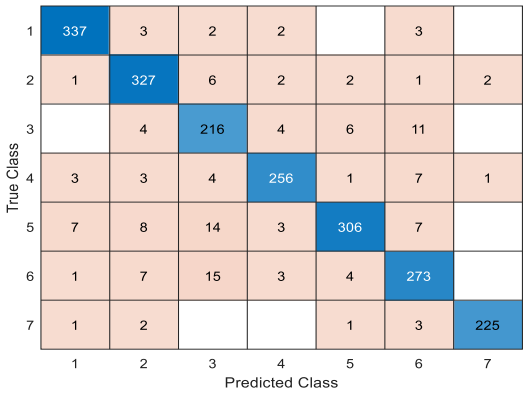


Figure 5. *Confusion matrix of the proposed CNN based method*

Table 3. *Overall performance of the proposed CNN model*

Metric	Value
Overall Accuracy	92.37%
Training Accuracy	96.8%
Validation Accuracy	93.5%
Top-1/2/3	92.37 / 96.79 / 98.66
Macro P/R/F1	0.9240 / 0.9239 / 0.9237
Weighted P/R/F1	0.9248 / 0.9237 / 0.9240
Cohen's Kappa	0.9107
Macro AUC	0.9949
Brier Score	0.0154
Number of Parameters	0.736 M
Throughput	142.2 FPS

In terms of overall performance, the CNN model achieves a test accuracy of 92.37%, which is the highest among all evaluated methods. Compared to the HOCv2 + Wavelet fusion approach (91.04%), this corresponds to an improvement of approximately 1.33%. Although this gain is relatively modest, it is important to note that the CNN does not rely on handcrafted feature design, but instead learns directly from raw data after STFT transformation. This highlights the strength of representation learning, enabling the model to automatically extract discriminative information from RF signals without the need for manual feature engineering.

The performance metrics reported in Table 3 indicate that the model not only achieves high accuracy but also produces reliable predictions. Specifically, the Macro F1-score reaches 0.9237, and the Weighted F1-score reaches 0.9240, reflecting stable performance across all UAV classes. Cohen's Kappa of 0.9107 demonstrates a very high level of agreement between predicted and true labels, after accounting for chance agreement. In addition, the Macro AUC of 0.9949 confirms the model's excellent class separability in the probability space. Notably, the Brier score is as low as 0.0154, indicating well-calibrated probability estimates, where the predicted confidence aligns closely with the true likelihood. It is worth noting that the performance gap among the training accuracy of approximately 96.8%, validation accuracy of approximately 93.5%, and testing accuracy of approximately 92.37% is relatively small. This observation suggests that the proposed model preserves stable generalization performance rather than simply memorizing the training samples. Furthermore, the Top-k accuracy results also show strong performance, with Top-1, Top-2, and Top-3 accuracies of 92.37%, 96.79%, and 98.66%, respectively. This suggests that the correct class is almost always included among the highest-probability predictions. To further examine the sensitivity of the proposed CNN to the training dataset size, it is important to emphasize that the network was intentionally designed as a lightweight architecture, containing only approximately 0.736 million trainable parameters. Moreover, several regularization and generalization enhancing strategies, including Batch Normalization, Dropout, Global Average Pooling, and validation based checkpoint selection, were employed to mitigate the risk of overfitting. The resulting high Macro AUC of 0.9949, Cohen's Kappa coefficient of 0.9107, and low Brier score of approximately 0.015 further indicate that the model achieves strong inter-class discrimination and produces well-calibrated predictions on unseen test samples.

A detailed class wise analysis presented in Table 4 shows that the CNN model achieves high and relatively uniform performance across all classes. The F1-scores for the classes are 0.953 (UAV-1), 0.947 (UAV-2), 0.873 (UAV-3), 0.922 (UAV-4), 0.913 (UAV-5), 0.874 (UAV-6), and 0.983 (UAV-7). It can be observed that all classes achieve F1-scores above 0.87, with UAV-7 reaching the highest value. More challenging classes such as UAV-3 and UAV-6 still maintain strong performance (approximately 0.87), indicating that the model generalizes well even for UAVs with similar RF fingerprints. Compared to the fusion based method in Section 3.3, the CNN further improves classification performance across most classes, particularly for UAV-1, UAV-2, and UAV-4, where time–frequency characteristics play a crucial role in device discrimination.

The confusion matrix in Figure 5 shows that the majority of samples are correctly classified, with strong concentrations along the diagonal. Compared to previous methods, the level of confusion between classes is significantly reduced, especially for UAV pairs with similar characteristics. Although some confusion still occurs in classes such as UAV-3 and UAV-6, the error rate is low and does not significantly affect overall performance.

The ROC curves in Figure 6 further demonstrate that all class specific curves closely approach the top left corner of the ROC space, indicating excellent discrimination capability of the proposed CNN framework. This observation is consistent with the high AUC values reported for all UAV classes, ranging from 0.9883 to 0.9998, and confirms the strong separability of the learned RF fingerprint representations in the probability space.

Table 4. *UAV classification performance of the proposed CNN model*

Class	Precision	Recall	F1-score
UAV-1	0.967	0.939	0.953
UAV-2	0.932	0.962	0.947
UAV-3	0.848	0.900	0.873
UAV-4	0.940	0.905	0.922
UAV-5	0.926	0.901	0.913
UAV-6	0.865	0.884	0.874
UAV-7	0.991	0.974	0.983

From a signal representation perspective, the superior performance of the CNN can be explained by its ability to simultaneously exploit multiple types of information embedded in the spectrogram. The STFT representation not only reflects the temporal evolution of spectral energy but also captures fine grained structures such as micro Doppler signatures and distortions induced by hardware characteristics. Convolutional layers in CNNs are capable of learning hierarchical time–frequency filters, allowing the model to automatically detect these discriminative patterns without manual feature engineering. This enables the model to capture complex relationships between time and frequency, which are difficult to represent using traditional features such as HOC or standalone wavelet methods.

Another important advantage of the model is its computational efficiency. With approximately 0.736 million parameters, the proposed CNN architecture is relatively lightweight compared to many modern deep learning models. At the same time, it achieves an inference speed of approximately 142 FPS, indicating strong potential for real time deployment. Compared to ensemble based approaches involving multiple models, the use of a single CNN significantly reduces system complexity while maintaining high performance. This is particularly important for practical applications in spectrum monitoring and UAV identification.

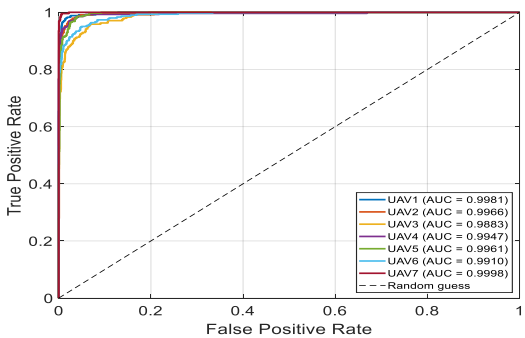


Figure 6. *ROC analysis of learned RF fingerprint representations for seven UAV classes*

Table 5. *Overall performance comparison of feature based methods*

Method	Test Accuracy (%)	Improvement vs HOC
HOC	43.27	-
HOCv2	83.87	+40.60
Wavelet	77.86	+34.59
HOCv2 + Wavelet	91.04	+47.77
Proposed CNN	**92.37**	**+49.10**

A comprehensive comparison in Table 5 reveals a clear trend of performance improvement when transitioning from HOC (43.27%) → HOCv2 (83.87%) → Wavelet (77.86%) → Fusion (91.04%) → CNN (92.37%). This progression reflects two key directions in enhancing RF fingerprinting performance: (i) improving feature quality by exploiting multiple representations (HOCv2, wavelet, fusion), and (ii) leveraging deep learning models to automatically learn feature representations directly from data. Among these, the CNN represents the second direction and demonstrates the highest performance without requiring handcrafted feature design. However, it is important to note that the performance gain between the CNN and the fusion based method is not substantial, indicating that well-designed handcrafted features can still achieve competitive performance. This observation opens up the possibility of combining both approaches, such as integrating handcrafted features into deep learning models or developing hybrid architectures, in order to leverage the strengths of both paradigms.

Although the CNN model provides the best classification performance, its interpretability remains a critical issue in RF spectrum monitoring and security oriented applications. To investigate the decision-making behavior of the proposed CNN, an occlusion sensitivity analysis was conducted on the input time–frequency spectrograms. Let X denote the input spectrogram and c denote the ground truth class label. The original prediction confidence is first computed as $p_c=P(y=c|X)$. A local occlusion window of size 16×16 is then systematically moved across the entire spectrogram. For each spatial location (i,j), the corresponding time–frequency patch $\Omega_{i,j}$ is replaced by the mean intensity value of the input image, yielding the occluded spectrogram $X_{\Omega_{i,j}}$. The importance score at location (i,j) is quantified by the resulting decrease in prediction confidence:

$$I(i,j) = p_c - P\left(y=c|X_{\Omega_{i,j}}\right). \tag{32}$$

The results in Table 6 show that occluding highly important regions causes a substantially larger decrease in classification confidence than occluding less important regions. On average, occluding important regions reduces the classification confidence by approximately 0.5467, whereas occluding less important regions leads to a much smaller decrease of about 0.2553. For UAV1 and UAV2, the confidence drops caused by occluding important regions reach 0.7475 and 0.7987, respectively, while occluding less important regions produces only negligible changes. These results indicate that the CNN does not make decisions based on random noise patterns or uniformly distributed global intensity features. Instead, it focuses on discriminative time–frequency structures that carry characteristic RF fingerprint information. Therefore, although the CNN is less directly interpretable than hand crafted features such as HOCv2 and Wavelet RF-DNA, the occlusion sensitivity analysis provides quantitative evidence supporting the reliability of the learned features used by the model.

Table 6. *Confidence-drop analysis using occlusion sensitivity for the CNN model*

UAV Class	Base confidence	After Important Mask	After Unimportant Mask	Important Drop	Unimportant Drop
UAV-1	0.8168	0.0693	0.7983	0.7475	0.0186
UAV-2	0.8611	0.0625	0.8617	0.7987	0.0000
UAV-3	0.3891	0.1107	0.1460	0.2784	0.2431
UAV-4	0.4711	0.0537	0.4158	0.4174	0.0553
UAV-5	0.7347	0.1531	0.0490	0.5817	0.6857
UAV-6	0.8540	0.2498	0.5071	0.6042	0.3469
UAV-7	0.4743	0.0753	0.0367	0.3990	0.4376
Mean	-	-	-	0.5467	0.2553

In summary, the results in this section demonstrate that the CNN model based on STFT spectrograms achieves the highest performance among all evaluated methods, while maintaining relatively low complexity and real time feasibility. These findings confirm that deep learning is an effective approach for UAV RF fingerprinting, particularly when combined with appropriate time–frequency representations.

4. Conclusion

This paper proposes a unified RF fingerprinting framework for device level UAV identification based on wideband radio signals. Unlike approaches that focus solely on a single feature domain or rely entirely on deep learning models, the proposed study develops a comprehensive pipeline that simultaneously exploits statistical features, multi resolution time–frequency features, and deep learning based representations within a unified dataset. This approach not only enables systematic comparison across different methods but also clarifies the roles and complementary characteristics of each representation domain in UAV RF fingerprinting.

Based on this framework, the study introduces the HOCv2 method to address the limitation of traditional HOC features in losing temporal dynamic information. Experimental results demonstrate that HOCv2 significantly improves classification performance, from 43.27% to 83.87%, confirming that preserving temporal variations is critical in dynamic UAV signal environments, where channel conditions continuously change. In addition, the Wavelet RF-DNA features effectively capture the multi scale spectral structure of RF signals, achieving an accuracy of 77.86%, and serving as an important complementary representation to statistical features.

Building on the complementary nature of these two feature groups, the study constructs a fused feature vector that combines HOCv2 and Wavelet RF-DNA, thereby improving the performance to 91.04%. These results confirm that jointly exploiting information from both the statistical domain and the time–frequency domain is an effective approach for enhancing the discriminative capability among UAV classes. However, handcrafted feature based methods still depend on feature design and parameter selection, which highlights the need for models capable of automatically learning feature representations directly from data.

To address this limitation, the paper proposes a lightweight CNN model operating on STFT spectrogram representations. The model achieves a high accuracy of 92.37%, along with strong evaluation metrics such as Macro AUC = 0.9949, Cohen's Kappa = 0.9107, and a low Brier score (0.0154), indicating not only accurate classification but also well-calibrated probability estimates. Notably, the model contains approximately 0.736 million parameters and achieves an inference speed of around 142 FPS, demonstrating its feasibility for real time UAV identification systems. Compared to more complex ensemble based approaches in previous studies, the use of a single CNN model provides an effective balance between performance and computational cost.

Based on the overall findings, two key conclusions can be drawn. First, in UAV identification tasks involving devices with subtle hardware differences and signals strongly affected by channel variations, feature design should prioritize preserving temporal dynamics rather than relying solely on global statistical characteristics. Second, different feature domains (statistical, time–frequency, and deep learning based representations) are not mutually exclusive but complementary, and integrating them within a unified framework is essential for achieving high and stable performance. Despite the promising results, the study still has several limitations. First, the experiments are conducted on data collected in a controlled environment (anechoic chamber), which does not fully reflect the complexity of real world channel conditions. Second, the model has not been evaluated under more challenging scenarios, such as burst-unseen, distance-unseen, or the presence of previously unseen UAV devices during inference. Third, although the CNN achieves strong performance, deeper integration between handcrafted features and deep learning (e.g., hybrid models) has not been fully explored. Furthermore, while the proposed CNN demonstrates strong classification performance under a closed set experimental setting, its robustness against realistic channel variations, environmental interference, and previously unseen UAV devices remains to be comprehensively investigated. Future work will therefore focus on extending the proposed framework through noise aware training, channel aware data augmentation, domain adaptation, and open set RF fingerprinting, to improve the robustness, scalability, and generalization capability of the system for real world deployment scenarios.

In future work, several important directions can be pursued. (i) Incorporating domain adaptation techniques and channel-invariant learning to improve generalization

under varying channel conditions. (ii) Exploring hybrid models that combine handcrafted features with deep learning to leverage the strengths of both approaches. (iii) Extending the framework to multimodal systems by integrating RF with radar or vision data, thereby enhancing the reliability of UAV identification systems in real world scenarios. In summary, this paper demonstrates that constructing a unified RF fingerprinting framework, which integrates handcrafted feature design and deep learning, can achieve high performance and practical deployability for UAV identification. These findings not only contribute methodologically but also provide a system-oriented perspective for future research in RF based device identification.

REFERENCES

[1] M. Ritchie, F. Fioranelli, and H. Griffiths, "Multistatic micro Doppler radar feature extraction for micro drone classification," *IET Radar, Sonar & Navigation*, vol. 11, no. 1, pp. 116–124, 2017.

[2] M. F. Al-Sa'd, A. Al-Ali, A. Mohamed, T. Khattab, and A. Erbad, "RF based drone detection and identification using deep learning approaches: An initiative towards a large open source drone database," *Future Generation Computer Systems*, vol. 100, pp. 86–97, 2019.

[3] R. Opromolla, G. Fasano, and D. Accardo, "A Vision based Approach to UAV Detection and Tracking in Cooperative Applications," *Sensors*, vol. 18, no. 10, Art. no. 3391, 2018.

[4] N. Soltanieh, Y. Norouzi, Y. Yang, and N. C. Karmakar, "A Review of Radio Frequency Fingerprinting Techniques," *IEEE Journal of Radio Frequency Identification*, vol. 4, no. 3, pp. 222–233, 2020.

[5] A. Jagannath, J. Jagannath, and P. S. P. V. Kumar, "A Comprehensive Survey on Radio Frequency (RF) Fingerprinting: Traditional Approaches, Deep Learning, and Open Challenges," *Computer Networks*, vol. 219, Art. no. 109455, 2022.

[6] N. Soltani, G. Reus-Muns, B. Salehi, J. Dy, S. Ioannidis, and K. R. Chowdhury, "RF Fingerprinting Unmanned Aerial Vehicles With Non Standard Transmitter Waveforms," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 12, pp. 15518–15531, 2020.

[7] T. J. O'Shea and J. Hoydis, "An Introduction to Deep Learning for the Physical Layer," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563–575, 2017.

[8] S. M. Kay, *Fundamentals of Statistical Signal Processing*. Upper Saddle River, NJ, USA: Prentice Hall, 1993.

[9] A. Swami and B. Sadler, "Hierarchical digital modulation classification using cumulants," *IEEE Transactions on Signal Processing*, vol. 48, no. 2, pp. 416–429, 2000.

[10] O. A. Dobre, Y. Bar-Ness, and W. Su, "Signal identification using RF-DNA fingerprints," *IEEE Transactions on Wireless Communications*, vol. 14, no. 5, pp. 2634–2645, 2015.

[11] X. Wang, Y. Zhang, H. Zhang, X. Wei, and G. Wang, "Identification and authentication for wireless transmission security based on RF-DNA fingerprint," *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, Art. no. 230, 2019.

[12] B. Hilburn, N. West, T. O'Shea, and T. Roy, "SigMF: The Signal Metadata Format," in *Proc. GNU Radio Conf.*, 2018.

[13] Y. Mo, H. Jianjun, G. Qian, Y. Hu, J. Zhang, and W. Yue, "Deep Learning Approach to UAV Detection and Classification by Using Compressively Sensed RF Signal," *Sensors*, vol. 22, no. 8, Art. no. 3072, 2022.

[14] H.-K. Le, V.-S. Doan, and V.-P. Hoang, "Ensemble of Convolution Neural Networks for Improving Automatic Modulation Classification Performance". *The University of Danang - Journal of Science and Technology*, vol. 20, no. 6.2, pp. 25–32, 2022, doi: 10.31130/ud-jst.2022.293E.