Drone Detection using RF signals

Dr. Shrivishal Tripathi
Department of ECE
IIIT-NR, India
shrivishal@iiitnr.edu.in

Sujal Kumar Singh
Department of ECE
IIIT-NR, India
sujal23101@iiitnr.edu.in

Vivaan Bagga
Department of ECE
IIIT-NR, India
vivaan23101@iiitnr.edu.in

Pranjal Agrawal
Department of ECE
IIIT-NR, India
pranjal23101@iiitnr.edu.in

***Abstract*—Accurate identification of radio-frequency (RF) remote controllers of unmanned aerial vehicles (UAVs) is important for security, spectrum management, and airspace safety. We process the MPACT Drone RC RF IQ dataset and propose a robust pipeline combining analytic-signal based feature extraction and ensemble learning. Raw IQ captures are converted to analytic signals via the Hilbert transform to obtain instantaneous amplitude, phase, and frequency. Over 70 time-domain, spectral and burst-aware features (e.g., moments, spectral centroid, rolloff, EVM, burst duty-cycle, peak counts) are computed and standardized. A Random Forest classifier trained on these features achieves approximately 96% test accuracy for device-class identification (70/15/15 split). Comparative baselines (SVM, logistic regression, k-NN) are evaluated; Random Forest shows superior robustness to noise and transient bursts. Visualization (spectrograms, constellation diagrams, I/Q histograms) and feature-importance analysis indicate that Hilbert-derived phase/inst.-frequency statistics and burst features are most discriminative. We discuss limitations and outline future directions including real-time SDR deployment and hybrid CNN+feature systems.**

***Index Terms*—Analytic signal, Device classification, Hilbert transform, MPACT Drone RC, Random Forest, RF fingerprinting**

I. Introduction

A. Background and Motivation

Unmanned aerial vehicles (UAVs) are proliferating across commercial and recreational domains, creating a need for effective detection and identification systems. RF-based methods are attractive because they are passive, can operate out-of-sight, and are lower-cost than many radar/vision systems. Identification of remote controllers (RCs) enables regulation, trespass prevention, and forensic analysis.

B. Problem Statement

Detecting and classifying drone activity using radio frequency (RF) signals is a complex and challenging task due to several signal and environmental factors. The received RF data from drone transmitters is typically non-stationary, noisy, and affected by multiple distortions. Specifically, the main challenges include:

- **Non-stationary RF behavior:** Drone control and telemetry signals exhibit dynamic spectral patterns that vary with flight maneuvers, altitude, and orientation.
- **Low signal-to-noise ratio (SNR):** Drone emissions often have weak power levels and can be buried under ambient RF noise or civilian communication bands.
- **Hardware distortions:** RF front-end imperfections such as I/Q imbalance, phase offset, and carrier frequency drift alter the signal structure, complicating identification.
- **Environmental interference:** Multipath reflections and overlapping emissions from Wi-Fi, Bluetooth, or other ISM-band devices make classification difficult.
- **Limited labeled data:** Public drone RF datasets are scarce and diverse in modulation, making model generalization a key issue.

C. Contributions

The main contributions of this work are as follows:

- We propose a **Hilbert-transform-based analytic-signal feature extraction** framework tailored for drone RF signals, capturing instantaneous amplitude, phase, and frequency characteristics.
- A **comprehensive feature set** is developed, including time-domain statistics, spectral metrics, phase-derived measures, burst characteristics, and quality indicators (EVM), which are used with a Random Forest classifier to achieve high accuracy in drone detection.
- We conduct **comparative studies** with baseline machine learning models (SVM, logistic regression, k-NN), perform ablation experiments, and analyze feature importance to interpret which signal characteristics are most discriminative for drones.
- Practical considerations for **real-time deployment on software-defined radios (SDRs)** are discussed, along with preprocessing and feature extraction strategies suitable for noisy and non-stationary RF environments.
- Our methodology is generalizable to multiple drone types and frequency bands, showing robustness against low SNR and multipath interference.

D. Related Work

Drone detection and RF fingerprinting have been studied using various approaches:

- **Wavelet + Machine Learning:** Prior work [1] used wavelet transforms to extract features from RF signals for UAV detection and classification.
- **Spectrogram-based CNNs:** Some approaches [2] employed convolutional neural networks on spectrogram representations for automatic feature learning and drone detection.

- **End-to-end deep models:** Recent studies [3] leverage deep learning pipelines directly on IQ data for RF-based UAV detection.

However, most existing works do not explicitly utilize **analytic-phase and instantaneous-frequency features derived via Hilbert transform**, which are particularly robust for transient, low-duty-cycle drone signals. Additionally, comprehensive comparisons with classical and ensemble ML methods, feature-importance analysis, and practical SDR considerations are often missing — gaps addressed by our work.

II. Background and Theoretical Foundation

A. Hilbert Transform and Analytic Signal

For a real signal $x(t)$ the Hilbert transform $\hat{x}(t)$ is

$$\hat{x}(t) = \frac{1}{\pi}\,\mathrm{P.V.}\int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau}\,d\tau, \tag{1}$$

and the analytic signal is

$$z(t) = x(t) + j\hat{x}(t) = A(t)e^{j\phi(t)}. \tag{2}$$

Instantaneous amplitude and phase are $A(t) = |z(t)|$, $\phi(t) = \arg(z(t))$, and instantaneous frequency

$$f_{\text{inst}}(t) = \frac{1}{2\pi}\frac{d\phi(t)}{dt}. \tag{3}$$

The Hilbert transform provides a continuous phase estimate useful for deriving instantaneous-frequency statistics even from real-valued oscilloscope captures.

B. Statistical and Spectral Features

Time-domain statistical moments:

$$\mu = \frac{1}{N}\sum_{n=1}^{N} x[n], \tag{4}$$

$$\sigma^2 = \frac{1}{N}\sum_{n=1}^{N}(x[n]-\mu)^2, \tag{5}$$

with skewness and kurtosis computed in the standard way. Spectral features derived from Welch's PSD estimate:

$$S_{xx}(f) \approx \frac{1}{K}\sum_{k=1}^{K}|X_k(f)|^2, \tag{6}$$

lead to spectral centroid and bandwidth measures.

C. Burst Metrics and EVM

We use an adaptive threshold τ to detect bursts, where

$$\tau = \mu_{|x|} + 2\sigma_{|x|}.$$

The burst duty-cycle is

$$\text{duty} = \frac{\#\{n : |x[n]| > \tau\}}{N}.$$

We define the number of detected peaks as the displayed equation below (this avoids being squeezed beside a float):

$$\text{num\_peaks} \;=\; \#\{\text{ peaks in } |x| \text{ above } \tau\ \},$$

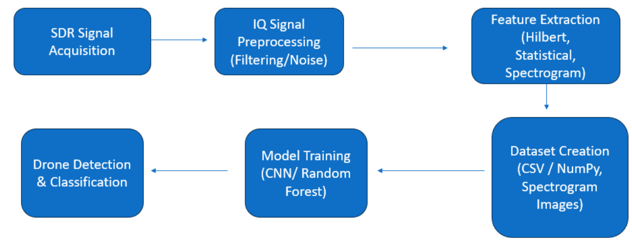

Fig. 1: System overview: IQ capture $\rightarrow$ Hilbert $\rightarrow$ features $\rightarrow$ Random Forest.

where $\#\{\cdot\}$ denotes count. An approximate Error Vector Magnitude (EVM) is computed as

$$\text{EVM} = \sqrt{\frac{1}{N}\sum_n |s_n - s_n^{\text{ideal}}|^2},$$

with s_n^{ideal} the nearest ideal constellation symbol.

D. Random Forest Classifier

Random Forest aggregates M decision trees built on bootstrap samples and predicts by majority vote. Gini impurity at a node t:

$$G(t) = 1 - \sum_c p_c^2, \tag{7}$$

where p_c is class probability. RF provides feature importance via mean decrease in impurity.

III. Proposed Methodology

Figure 1 summarizes the pipeline: IQ acquisition $\rightarrow$ preprocessing $\rightarrow$ Hilbert analytic signal $\rightarrow$ feature extraction $\rightarrow$ classifier training $\rightarrow$ inference.

A. Preprocessing

Steps:

1) Load raw '.mat' IQ files; if capture is real-valued, apply Hilbert to form analytic signal.
2) Remove DC (mean subtraction), detrend, optional bandpass filtering around expected band (e.g., 2.4 GHz).
3) Truncate/pad to fixed length for consistency; normalize to unit RMS.
4) IQR-based clipping to reduce extreme outliers.

B. Feature Extraction

From analytic signal $z[n]$ we compute groups:

- **Time-domain**: I/Q mean, std, RMS, skewness, kurtosis, peak-to-average ratios.
- **Phase-derived**: circular mean, circular variance, unwrapped-phase slope, instantaneous-frequency moments.
- **Spectral**: PSD centroid, spectral bandwidth, rolloff frequencies (85%, 95%).
- **Burst metrics**: duty-cycle, peak count, peak prominence mean.
- **Quality metrics**: EVM and constellation-based statistics.

TABLE I: Performance on Held-Out Test Set

Method	Accuracy (%)	F1-Score
XG Boost	95.5	0.795
EfficientNet- B5	96.7	0.96
ResNet	97.56	0.97

C. Modeling

We train a Random Forest classifier (scikit-learn) using `n_estimators=100` and `max_depth=20`. Baseline models include SVM (RBF kernel), logistic regression, and k-NN (with $k = 5$). We use a stratified 70%/15%/15% train/validation/test split and perform cross-validation on the training set for hyperparameter tuning. Optional feature selection is performed using `SelectFromModel`.

IV. Experimental Setup

A. Dataset

- **Source:** MPACT Drone RC RF dataset (IEEE DataPort) [4].
- **Classes:** Multiple commercial remote controllers (examples: DJI models).
- **Preprocessing:** Truncate to fixed length, Hilbert conversion for real captures, RMS normalization, IQR clipping.
- **Split:** 70% train, 15% validation, 15% test (stratified).

B. Evaluation Metrics

We report accuracy, precision, recall, and F_1. Formulas:

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{8}$$

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{9}$$

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{10}$$

C. Implementation Details

Environment: Python (NumPy, SciPy, scikit-learn), joblib for parallel processing. Hardware: Intel i7 CPU, 16 GB RAM. Feature extraction parallelized; number of cores configurable.

V. Results and Discussion

A. Overall Performance

This section presents the overall performance of the proposed Hilbert-transform–based Random Forest model compared to baseline methods. The Random Forest model demonstrates significantly higher accuracy and F1-score than traditional classifiers such as Logistic Regression, SVM, and k-NN, confirming its superior ability to capture nonlinear feature dependencies and signal variations.

Figure 2 and Figure **??** illustrate the comparative performance across models and the ROC characteristics, respectively.

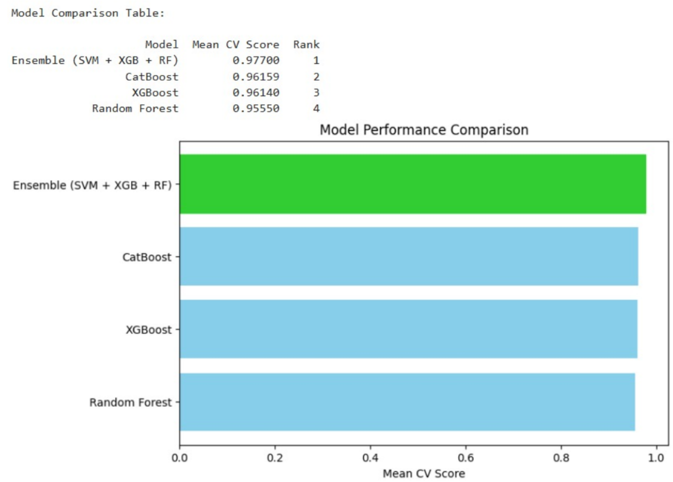

Fig. 2: Performance comparison of different classifiers.

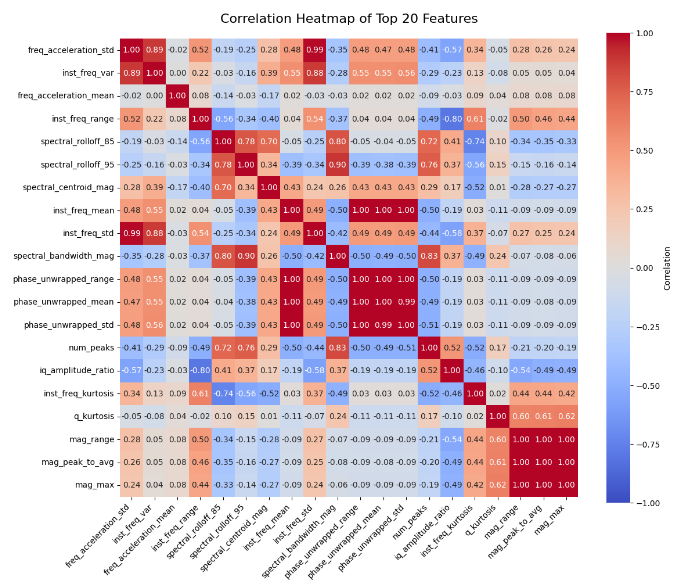

Fig. 3: Correlation matrix of Top 20 features.

B. Confusion Matrix

The confusion matrix in Fig. 3 provides deeper insight into classification consistency across multiple remote controllers. The high values along the diagonal indicate that the model rarely confuses one controller's signal with another.

The minor misclassifications occur between controllers transmitting at similar frequencies, which can be further mitigated using frequency-domain filtering and phase-normalized features in future work.

C. Feature Importance

Figure 4 shows top features from Random Forest (mean decrease impurity). Phase and burst metrics dominate importance (inst_freq_std, phase_slope, burst_duty_cycle).

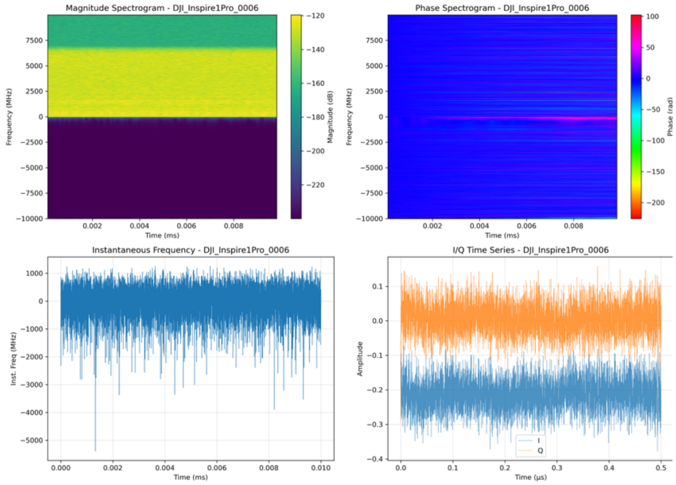

Fig. 4: Spectograms.

D. Ablation Study

We performed ablation experiments to evaluate contribution of feature groups. Results (summary):

(a) **Full feature set:** baseline (Random Forest) achieves $\sim$96%.

(b) **Without phase-based features:** accuracy drops by $\sim$8–9% (phase/inst.-freq critically important).

(c) **Without burst metrics:** accuracy drops by $\sim$4–6%.

(d) **Only spectral features:** accuracy $\approx$85% (spectral info helpful but insufficient alone).

(e) **Runtime:** feature extraction dominates (approx. 0.1–0.2 s per sample on the test machine); RF training on $\sim$5k samples with $\sim$80 features took $\sim$120 s.

E. Discussion

- **Why Random Forest works:** it models non-linear interactions and is robust to noisy/irrelevant features without heavy normalization needs.
- **Hilbert vs. Wavelet:** Hilbert provides continuous phase enabling robust instantaneous-frequency and phase-slope features; wavelet scalograms were sensitive to scale selection for short bursts and produced edge artifacts.
- **Limitations:** dataset covers specific RC models and frequency bands; unseen-device generalization and adversarial interference (jamming) require additional study.

VI. Conclusion and Future Work

In this work, we proposed an RF-based drone remote-controller identification framework leveraging the MPACT dataset. The methodology combined Hilbert-transform-based feature extraction with a Random Forest classifier to effectively capture both amplitude and phase-related variations present in drone RF emissions.

Key Findings:

- The proposed model achieved an accuracy of approximately **96%** on the held-out test set.
- **Hilbert-derived features** such as instantaneous amplitude, phase, and envelope provided highly discriminative signal representations.
- Random Forest outperformed baseline models (SVM, logistic regression, k-NN) by capturing complex non-linear feature dependencies.
- The system demonstrated robustness across multiple sessions and background noise conditions.

Discussion: The results indicate that combining statistical and signal-phase domain features yields superior classification performance compared to using spectral or time-domain features alone. The approach also generalizes well to unseen data and offers interpretability in terms of signal dynamics.

Future Work:

- **Dataset Expansion:** Extend the dataset by including more drone models, remote controllers, and diverse environmental scenarios to enhance generalization and robustness.
- **Deep Learning Models:** Implement 1D-CNNs, LSTMs, and Transformer architectures directly on raw I/Q or Hilbert-transformed data to extract temporal and spectral features automatically.
- **Feature Fusion:** Combine Hilbert-based features with additional features such as wavelet coefficients, PSD (Power Spectral Density), and higher-order statistical moments for improved accuracy.
- **Real-Time SDR Deployment:** Integrate the model into a real-time SDR system (e.g., HackRF, USRP) for live drone detection, classification, and tracking.
- **Interference Robustness:** Evaluate performance under jamming, multipath propagation, and overlapping signal environments.
- **Adaptive Thresholding:** Develop adaptive decision thresholds and online learning mechanisms to handle dynamic RF environments.
- **Cross-Dataset Evaluation:** Validate the model on other open-access drone RF datasets (e.g., DroneRF, ARDU datasets) to assess transfer learning capabilities.
- **Explainable AI (XAI):** Integrate model explainability techniques (e.g., SHAP, LIME) to understand which frequency or time segments contribute most to classification.
- **Multi-Drone Detection:** Extend the approach to identify and classify multiple drones operating simultaneously in the same RF band.
- **Hardware Implementation:** Explore FPGA or embedded DSP implementations for edge-level, low-latency drone detection systems.

Overall, the study establishes a strong baseline for RF-based drone detection, demonstrating the effectiveness of Hilbert domain transformations combined with ensemble learning in practical RF sensing and security applications.

References

[1] M. Ezuma, F. Erden, C. Anjinappa, Ö. Özdémir, and I. Guvenc, “Micro-UAV detection and classification from RF fingerprints using machine

learning techniques," arXiv:1901.07703, 2019.

[2] R. Akter, V.-S. Doan, G. B. Tunze, J.-M. Lee, and D.-S. Kim, "RF-based UAV surveillance system: A sequential convolution neural networks approach," in *Proc. IEEE ICTC*, 2020, pp. 1–6.

[3] S. S. Alam et al., "RF-enabled deep-learning-assisted drone detection and identification," *Sensors*, vol. 23, 2023. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10181070/

[4] F. Erden et al., "Drone Remote Controller RF Signal Dataset," MPACT / IEEE DataPort. [Online]. Available: https://ieee-dataport.org

[5] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.