Acoustic Drone Detection and Classification

Statistical Signal Processing and ML Approaches

İhsan Mert Muhacıroğlu-504231345 Yasin Apalan-504241334

Statistical Signal Processing – Term Project

Istanbul Technical University
Department of Telecommunication Engineering

Outline

Introduction

Problem Definition

The Challenge

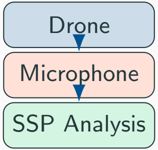
- UAVs widespread $\Rightarrow$ security & privacy risks
- Detection modalities: radar, RF, vision, acoustics
- **Goal:** Detect & classify drones via acoustic signatures

Why Acoustic?

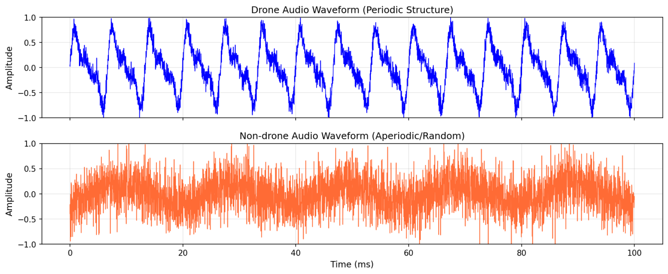
- Passive sensing (no emission)
- Low-cost hardware
- Distinct periodic structure
- SSP provides interpretable features

Dataset
Al-Emadi et al., “Audio Based Drone Detection...”

`github.com/saraalemadi/DroneAudioDataset`



Audio Waveform: Drone vs Non-drone



Top: Drone signal shows regular oscillations (propeller).
Bottom: Non-drone lacks pattern.

Key Insight: Drone signals exhibit clear periodicity from propeller rotation; non-drone sounds are aperiodic.

Drone Acoustic Characteristics

Physical Properties

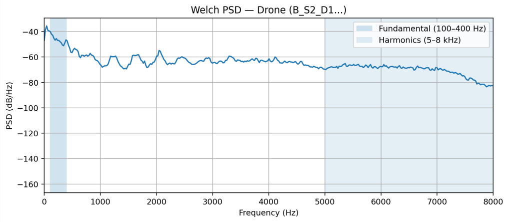
- Propeller rotation $\Rightarrow$ periodic components
- Fundamental: $f_0 \approx 100$–$400\,\mathrm{Hz}$
- Harmonics: up to 5–8 kHz
- Short-term quasi-stationarity

Frequency Formula

$$f_0 = \frac{\mathrm{RPM} \times N_b}{60}$$

f_0: fundamental freq, N_b: blade count

Ex: 6000 RPM, 2 blades $\Rightarrow f_0 = 200\,\mathrm{Hz}$



PSD: strong peak at f_0 with harmonics at $2f_0, 3f_0, \ldots$

Mathematical Signal Model

Harmonic + Noise Model

$$x(t) = \underbrace{\sum_{k=1}^{K} A_k \cos(2\pi k f_0 t + \phi_k)}_{\text{Deterministic (propeller)}} + \underbrace{n(t)}_{\text{Stochastic}} \tag{1}$$

$x(t)$: signal A_k: k-th harmonic amplitude f_0: fundamental ϕ_k: phase K: # harmonics $n(t)$: noise

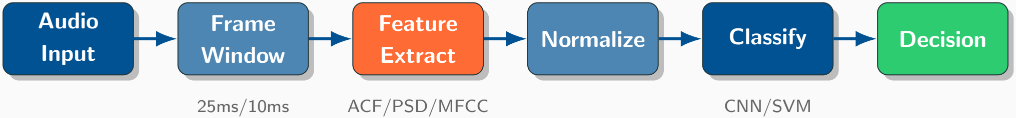
Parameters:

- f_0: fundamental frequency (RPM-dependent)
- A_k, ϕ_k: harmonic amplitude/phase
- $K \sim 10$–20 harmonics
- $n(t)$: environmental noise (AWGN)

SSP Implications:

- Periodicity $\Rightarrow$ ACF peaks
- Harmonics $\Rightarrow$ PSD lines
- Time-variation $\Rightarrow$ STFT
- Spectral shape $\Rightarrow$ MFCC

Processing Pipeline



Preprocessing:

- Frame: 25 ms (quasi-stationary)
- Hop: 10 ms (60% overlap)
- Window: Hamming

Features:

- ACF: Periodicity detection
- PSD: Frequency content
- MFCC: Spectral envelope

Signal Processing Methods

Autocorrelation Function (ACF)

Definition

$$R_{xx}(\tau) = \mathbb{E}[x(t)\,x(t+\tau)]$$

R_{xx}: autocorrelation τ: lag $\mathbb{E}$: expectation

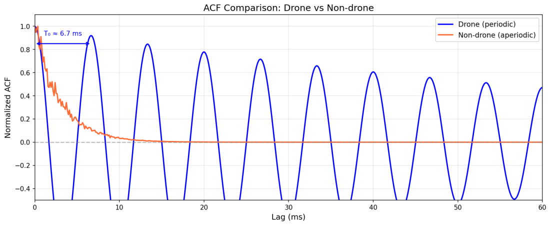
Discrete Normalized ACF

$$r_{xx}[k] = \frac{\sum_{n=0}^{N-1-k} x[n]\,x[n+k]}{\sum_{n=0}^{N-1} x^2[n]}$$

$r_{xx}[k]$: normalized ACF at lag k N: length

Properties:

- $r_{xx}[0] = 1$ (max at zero lag)
- Periodic $\Rightarrow$ peaks at $k = mT_0$
- Noise $\Rightarrow$ fast decay



Blue: Drone ACF - periodic peaks at $T_0 \approx 6.7$ms.

Orange: Non-drone - rapid decay.

Power Spectral Density (PSD)

Wiener-Khinchin Theorem

$$S_{xx}(f) = \mathcal{F}\{R_{xx}(\tau)\}$$

S_{xx}: PSD $\quad \mathcal{F}$: Fourier transform $\quad R_{xx}$: ACF

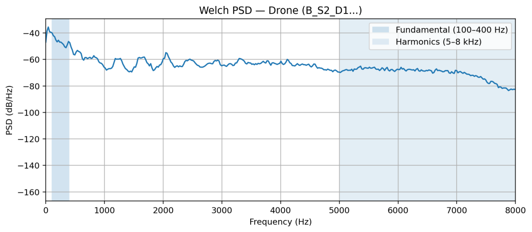
Welch Estimator

$$\hat{S}_{xx}(f) = \frac{1}{KU} \sum_{i=0}^{K-1} |X_i(f)|^2$$

K: # segments $\quad U$: window energy $\quad X_i$: segment FFT

Advantages:

- Variance reduction via averaging
- Reduced spectral leakage



Drone PSD: Fundamental f_0 (100-400Hz) + harmonics to 5-8kHz.

Key Insight: Harmonics appear as peaks at $f_0, 2f_0, 3f_0, \ldots$

Short-Time Fourier Transform (STFT)

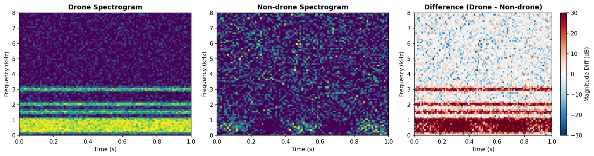
Time-Frequency Analysis

$$X[m,k] = \sum_{n=0}^{N-1} x[n]\, w[n-mH]\, e^{-j2\pi kn/N} \quad \Rightarrow \quad \mathcal{S}[m,k] = |X[m,k]|^2$$

$X[m,k]$: STFT (frame m, freq k) w: window H: hop N: FFT size $\mathcal{S}$: spectrogram

Parameters:

- Window: 25 ms Hamming
- FFT: 512 points
- Hop: 10 ms



Left: Drone (bands). Mid: Non-drone. Right: Difference shows harmonics.

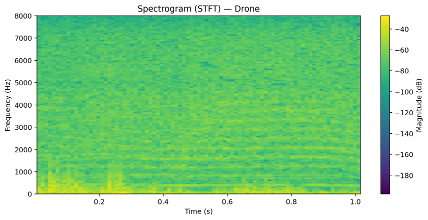
Uncertainty Principle

$$\Delta t \cdot \Delta f \geq \frac{1}{4\pi}$$

Short window $\Rightarrow$ good time, poor freq

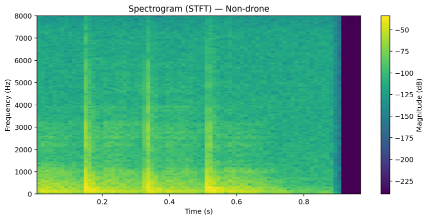
Spectrogram: Drone vs Non-drone

Drone



Stable horizontal bands = persistent harmonics

Non-drone



No consistent structure, random energy

Key Insight: Drone spectrograms show persistent horizontal bands; non-drone signals lack this pattern.

MFCC: Mel-Frequency Cepstral Coefficients

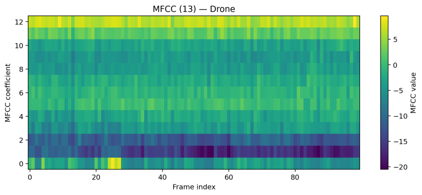
Processing Steps:

1. Pre-emphasis: $y[n] = x[n] - 0.97\,x[n-1]$
2. Framing: 25 ms, 10 ms hop
3. Windowing: Hamming
4. FFT $\rightarrow$ Power spectrum
5. Mel filterbank (40 filters)
6. Log compression
7. DCT $\rightarrow$ Cepstral coefficients

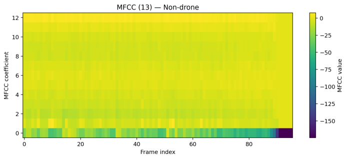
Mel Scale

$$m = 2595 \log_{10}\left(1 + \frac{f}{700}\right)$$

m: mel freq f: Hz. Mimics human hearing.



MFCC heatmap: x=time, y=13 coefficients. Drone pattern.



Non-drone: less structured, random variation.

MFCC: DCT and Delta Features

DCT Formula

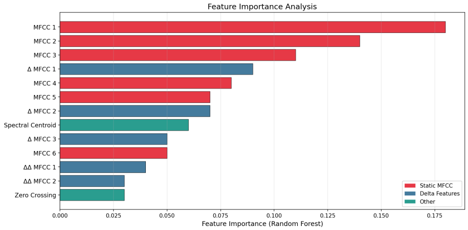
$$c_n = \sum_{m=1}^{M} \log(E_m) \cos\left[\frac{\pi n}{M}\left(m - \frac{1}{2}\right)\right]$$

c_n: n-th coefficient E_m: mel filter energy M: 40 filters

Feature Vector:

- Static MFCC: 13 (c_0–c_{12})
- Δ (velocity): 13
- $\Delta\Delta$ (acceleration): 13

Total
$13 + 13 + 13 = \mathbf{39}$ features/frame



Random Forest ranking: MFCC 1-3 most important, then Δ features.

Feature Summary

Feature	Domain	Drone Signature	Interpretation
ACF	Time	Peaks at T_0	Periodicity strength
PSD	Frequency	Lines at kf_0	Harmonic content
STFT	Time-Freq	Horizontal bands	Time evolution
MFCC	Cepstral	Distinct envelope	Compact spectral shape

Best for SVM/RF
MFCC statistics, PSD band energies, ACF peak features

Best for Deep Learning
CNN: Mel-spectrogram, LSTM: MFCC sequences, CRNN: Combined

Classification

Classification Models

Traditional ML
SVM (RBF kernel):

- Input: Feature statistics
- $K(\mathbf{x}, \mathbf{y}) = \exp\left(-\gamma \|\mathbf{x} - \mathbf{y}\|^2\right)$

γ: kernel width $\mathbf{x}, \mathbf{y}$: feature vectors

Random Forest:

- Ensemble of decision trees
- Feature importance ranking

Deep Learning
CNN: Best accuracy 96.2%

- Input: Mel-spectrogram (2D)
- Learns spectro-temporal patterns

LSTM: MFCC sequences, temporal deps.

CRNN: CNN + LSTM hybrid

Dataset & Experimental Setup

DroneAudioDataset

- Sampling: 44.1 kHz, 1 sec clips
- Environment: Indoor
- Augmentation: Noise added

Classes:

- **Binary:** Drone / Non-drone
- **Multi:** Bebop, Mambo, Phantom, Unknown

Training

- Split: 70/15/15%
- Optimizer: Adam (LR=0.001)
- Early stopping: patience=10

Metrics
Accuracy, F1, AUC-ROC, Confusion matrix, SNR robustness

Results

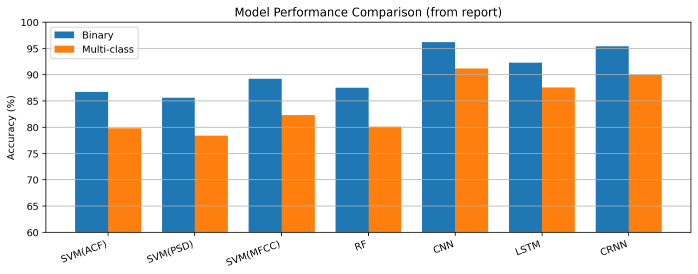
Performance Comparison

Model	Binary Acc.	Multi Acc.	F1	AUC
SVM (ACF)	86.7%	79.8%	0.861	0.894
SVM (PSD)	85.6%	78.4%	0.849	0.878
SVM (MFCC)	89.2%	82.4%	0.887	0.912
Random Forest	87.5%	80.1%	0.871	0.901
LSTM	92.3%	87.6%	0.919	0.951
CRNN	95.4%	90.1%	0.951	0.978
CNN	**96.2%**	**91.2%**	**0.958**	**0.981**

Binary: Drone vs Non-drone. Multi: 4-class. F1/AUC for binary task.

Key Insight: Deep learning outperforms traditional ML by +7%; CNN achieves best overall.

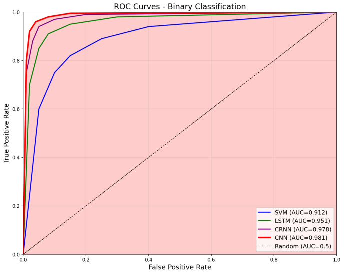
Results: Visual Comparison



Blue: Binary classification. Orange: Multi-class. CNN leads both; DL consistently outperforms ML.

■ Binary ■ Multi-class

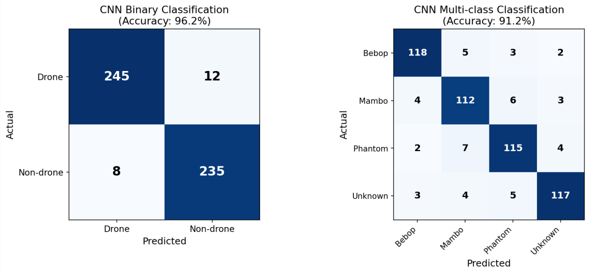
ROC Curves



ROC = TPR vs FPR. Top-left = better. AUC values in legend.

Key Insight: CNN achieves AUC = 0.981, significantly outperforming SVM (0.912).

Confusion Matrices (CNN)

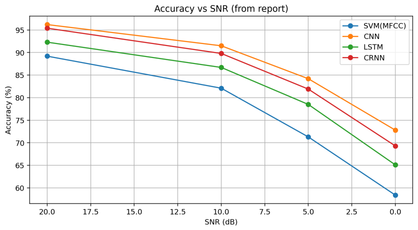


Left: Binary (245 TP, 235 TN). Right: Multi-class shows some Mambo-Phantom confusion.

Binary: Precision 95.3%, Recall 96.8%

Multi: Diagonal dominance = good

Noise Robustness (SNR Analysis)



X: SNR level (20dB=clean, 0dB=equal noise). All degrade, CNN slowest.

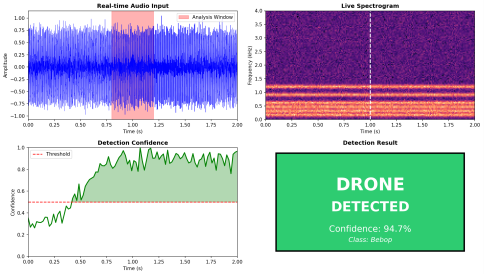
Observations:

- All models degrade with noise
- CNN most robust
- CRNN competitive at low SNR

At 0 dB SNR:

- CNN: **72.8%**
- CRNN: 69.3%
- LSTM: 64.1%
- SVM: 58.4%

Real-time Detection Concept



Top-left: Waveform + window. Top-right: Live spectrogram. Bottom: Confidence & result.

Key Insight: System processes audio in real-time (<10ms on GPU), continuous detection.

Conclusion

Discussion

SSP Feature Analysis

- **ACF:** Periodicity, lacks spectral detail
- **PSD:** Harmonics, loses temporal info
- **MFCC:** Best balance

Why CNN Wins?

- Local spectro-temporal patterns
- Translation invariance
- Hierarchical learning

Limitations

- Indoor dataset only
- Low SNR ($<$5 dB) challenging
- Only 3 drone types

Computational Cost

- SVM: $<$1 ms
- CNN: 5–10 ms (GPU)
- Real-time feasible

Conclusion & Future Work

Conclusions

1. Acoustic detection: practical & low-cost
2. SSP provides interpretable features
3. Best feature: MFCC
4. Best model: CNN (96.2%)
5. DL improves +7% over ML
6. Robust to 5 dB SNR

Future Work

1. Outdoor experiments
2. Edge deployment (RPi)
3. RF + acoustic fusion
4. Transformer architectures
5. Domain adaptation
6. Array processing (DOA)

Thank You!

Questions & Discussion

İhsan Mert Muhacıroğlu & Yasin Apalan
Istanbul Technical University
muhaciroglu19@itu.edu.tr & apalan18@itu.edu.tr