Drone Classification from RF Signals: A Comparative Study of Convolutional Networks and Attention Mechanisms

Luis Paulo Albuquerque Guedes[1], Pedro Henrique Monteiro Guedes[1] e Rêmulo Caminha[1]
[1]Marinha do Brasil, Rio de Janeiro/RJ - Brasil

Abstract— **The classification of unmanned aerial vehicles (UAVs) via radio-frequency (RF) signals employs advanced signal-processing and machine-learning techniques to identify and categorize emissions, playing a fundamental role in security and surveillance applications in sensitive environments. The present study conducts a comparative analysis between the VGG-16 and Transformer architectures, aiming to identify preprocessing and model configurations that maximize classification accuracy for drone RF signals without compromising computational feasibility in defense-embedded systems. Applying the VGG-16 model with 20 ms time blocks resulted in approximately 97 % accuracy and F1-score, outperforming classical methods (linear regression and k-NN) by up to 17 percentage points. Furthermore, it was found that all deep models exhibited significant gains when operating on spectrogram inputs, substantially surpassing traditional approaches.**

Keywords— **RF-based UAV Classification, CNN, Vision Transformers.**

I. Introduction

The classification of drones by radiofrequency (RF) is based on the use of advanced signal processing and machine learning techniques to identify and categorize RF emissions, playing a critical role in security and surveillance. Its historical development includes tests of radio-controlled balloons in 1917, the Hs 293 missile in 1935, the V-1 bombs in 1944, the MQ-1 Predator in 1950, the Zenit-2 spy satellite in 1964, the Matra MILAN combat drone in 1970, the use of UAVs in Lebanon in 1982, and its widespread adoption in the Afghanistan War in 2001 [1].

In current conflicts—such as in Ukraine and the Middle East—small drones have exposed vulnerabilities in defense systems, leading the U.S. Army to plan over US$ 400 million in 2025 for the development of integrated counter-drone systems [2]. Beyond the military context, UAVs threaten critical infrastructure through targeted attacks, smuggling, espionage, and collisions, which has driven proposals for multisensor fusion, artificial intelligence, and advanced machine learning algorithms to identify drone models and predict their trajectories and intentions.

II. Problem Formulation

Due to the increasing deployment of UAVs in civilian and military operations, there is a need for precise and efficient methods to classify drones based on radiofrequency (RF) signals. The key issues to be addressed include:

- **Temporal Segmentation**: How to determine the optimal granularity (30, 20, 15, 10, or 5 ms) for generating RF spectrograms that preserve discriminative information without incurring computational overload.
- **Spectrogram vs. Block Trade-off**: What is the impact of the number of spectrograms generated per unit time (i.e., processing block size) on accuracy and inference cost for the evaluated models.
- **Architecture Comparison**: In which scenarios attention-based models (Vision Transformers) outperform or fall short of convolutional architectures (VGG-16) in the task of classifying drone RF signals.

The objective of this work is to formulate, implement, and experimentally evaluate these questions by conducting a comparative analysis between VGG-16 and Transformers to identify preprocessing and architecture configurations that maximize classification accuracy for drone RF signals while maintaining computational feasibility for defense-embedded systems.

III. Related Work

RF-based drone classification is an essential research field, as it allows the identification and categorization of unmanned aerial vehicles from their RF emissions [3]. This capability proves particularly critical in security and surveillance applications in sensitive areas. To improve the accuracy and efficiency of these classifications, advanced signal processing and machine learning techniques have been employed.

Among the processing approaches, the *Short-Time Fourier Transform* (STFT) stands out for converting RF signals to the time-frequency domain, enabling the extraction of discriminative features. In [3], STFT was used to generate time-frequency spectra encoded as 2D images, which were fed into a CNN and achieved high performance on datasets such as DroneRF and DroneRFa. Complementarily, spectrogram analysis has been used to capture the evolution of frequency signatures over time, extracting statistics such as mean, variance, and spectral entropy, which are then classified by ensemble methods such as XGBoost and KNN [4].

In the deep learning domain, *Convolutional Neural Networks* enable the recognition of subtle patterns in RF signal representations. Residual CNN models have been applied in multipath scenarios, demonstrating robustness even under challenging conditions [5]. Additionally, object detection algorithms like YOLO have been adapted to treat RF spectrograms as images, improving performance through strategic annotation of transmission bursts [6].

Luis Paulo A. Guedes, luis.albuquerque@marinha.mil.br; Pedro H. M. Guedes, pedrohmguedes@gmail.com; Rêmulo Caminha, remulocaminh@gmail.com.

However, relevant challenges persist. The scarcity of comprehensive datasets limits the generalization capability of models [7], and the careful selection of features extracted from the signals—such as mean, skewness, and entropy—is fundamental to maximizing classification effectiveness [8]. Emerging techniques, such as sigmoid calibration, have been integrated into frameworks to adjust predicted probabilities and increase reliability in multiclass tasks [9]. Meanwhile, proposals that unify detection and classification in a single pipeline, as demonstrated in [10], offer efficiency gains in scenarios with multiple simultaneous signals.

IV. Materials and Methods

The detection system involves a UAV, its controller device, and two receiving stations responsible for measuring the signal strength emitted by the UAV—the first (Rx1) dedicated to the low-frequency band and the second (Rx2) to the high-frequency band—as illustrated in Figure 1. These RF transmissions correspond to the communication link between the UAV and the controller, being captured by environment-specific receivers, whose characteristics are detailed in Table 3. Finally, the collected data are transferred, stored, and processed on a computer or other interface equipment capable of handling and presenting RF data [11].

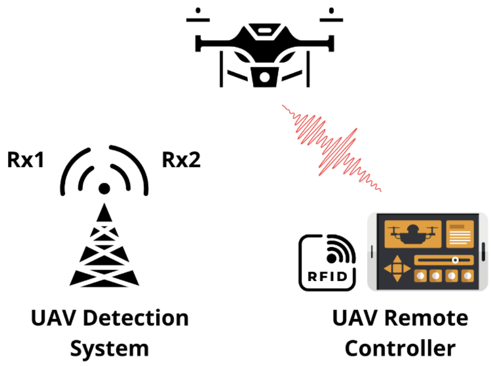

Fig. 1: Schematic of the detection system.

The DroneDetect corpus [12] is a dataset of RF signals from unmanned aerial systems, collected via BladeRF SDR and processed in GNURadio, designed for machine learning–based detection and classification tasks. It includes I/Q recordings of seven UAV models—AIR, MA1, MAV, INS, MIN, PHA, and DIS—sampled at 60 MHz with a 28 MHz bandwidth. In this study, the signals were segmented into blocks of 5, 10, 15, 20, 25, and 30 ms, normalized using the z-score, and subjected to STFT (1024-point window, 120-sample overlap). The magnitudes in dB of the positive frequencies were converted into 224x224 images (inferno colormap).

A grid search algorithm was implemented to find the best hyperparameters for two spectrogram-based drone classification architectures:

- **VGG-16**: consists of a `Flatten` block, followed by a fully connected layer with 256 ReLU units, a 50% `Dropout`, and a softmax classification layer. Optimal hyperparameters: learning rate 1×10^{-5}, 256 dense units, dropout rate 0.5.
- **ViT-Tiny**: uses 32x32 patches projected to 32-dimensional embeddings, followed by four Transformer blocks with two attention heads and a 64-dimensional MLP, before global pooling and a softmax head. Optimal hyperparameters: learning rate 1×10^{-5}, embedding dimension 32, MLP dimension 64, dropout rate 0.3.

Both models used 224x224 images, a batch size of 32, and a fixed seed of 42. The data were split stratified: 80% training, 10% validation, 10% testing. During training, data augmentations (rescaling, rotations up to $\pm 8^\circ$, shifts up to 5%, horizontal flips) were applied. The `Adam` optimizer was used with initial learning rate 1×10^{-5}, and cross-entropy loss. Callbacks saved checkpoints each epoch and performed early stopping with a patience of 10 epochs.

While VGG-16 exploits local patterns via deep convolutions, ViT-Tiny models global dependencies through attention mechanisms.

The dataset contains no background noise, so all simulations considered exclusively drone signals without ambient noise or interference.

In the first simulation, a block size of 50 ms was employed (Table I). Subsequent simulations were performed with 30 ms (Table II), 20 ms (Table III), 15 ms (Table IV), 10 ms (Table V), and finally 5 ms (Table VI), with the corresponding confusion matrices detailed in 2, 3, 4, 5, 6, and 7, respectively. Each block duration configuration allowed the investigation of the impact of temporal resolution on classification accuracy for the VGG-16 and Transformer architectures.

V. Results and Analysis

A. Average Performance Analysis by Drone Type across Models

Overall, when averaging recall for each drone type across all block durations:

- **VGG-16 vs. ViT-Tiny**: VGG-16 consistently outperforms the Transformer in almost all classes, with the largest gains in "AIR" and "MP2."
- **Class `MIN`**: Both models achieve nearly 100% accuracy on "MIN," suggesting this drone type is spectrally distinct from the others.
- **Classes `DIS` and `PHA`**: High average accuracy (> 0.95 for VGG-16 and ≈ 0.94 for the Transformer), indicating clear spectral characteristics.
- **Most challenging classes**: "MP1" and especially "MP2," where the Transformer drops to ~ 0.80, while VGG-16 remains around ~ 0.95.

Practical Conclusion: As shown in Figures 8, 9 and 10, the VGG-16 (CNN) demonstrates greater robustness and consistency in classifying RF signals of these drones at different temporal resolutions (block sizes), particularly in more ambiguous cases ("AIR," "MP2"). The Transformer, although competitive for some classes (e.g., "MIN," "INS"), struggles with spectral overlap for drones with similar signatures ("AIR" vs. "MP2").

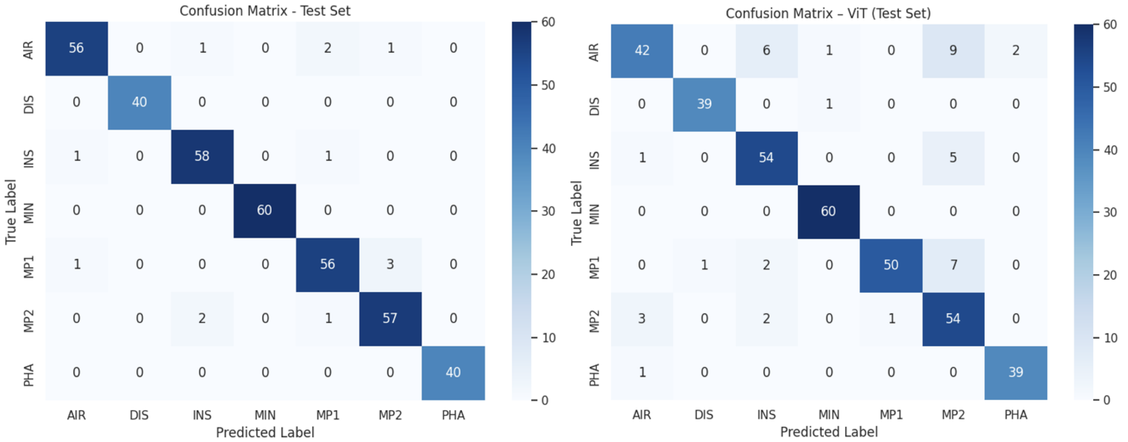

Fig. 2: Confusion Matrix - Results for VGG-16 and Transformer (Block = 50 ms)

TABLE I: Comparison of Classification Results for VGG-16 and Transformer (Block = 50 ms)

	VGG-16					Transformer			
Class	**Precision**	**Recall**	**F1-score**	**Support**	**Class**	**Precision**	**Recall**	**F1-score**	**Support**
AIR	0.9655	0.9333	0.9492	60	AIR	0.8936	0.7000	0.7850	60
DIS	1.0000	1.0000	1.0000	40	DIS	0.9750	0.9750	0.9750	40
INS	0.9508	0.9667	0.9587	60	INS	0.8438	0.9000	0.8710	60
MIN	1.0000	1.0000	1.0000	60	MIN	0.9677	1.0000	0.9836	60
MP1	0.9333	0.9333	0.9333	60	MP1	0.9804	0.8333	0.9009	60
MP2	0.9344	0.9500	0.9421	60	MP2	0.7200	0.9000	0.8000	60
PHA	1.0000	1.0000	1.0000	40	PHA	0.9512	0.9750	0.9630	40
Accuracy		0.9658		380	**Accuracy**		0.8895		380
Macro avg	0.9692	0.9690	0.9690	380	**Macro avg**	0.9045	0.8976	0.8969	380
Weighted avg	0.9659	0.9658	0.9658	380	**Weighted avg**	0.8984	0.8895	0.8893	380

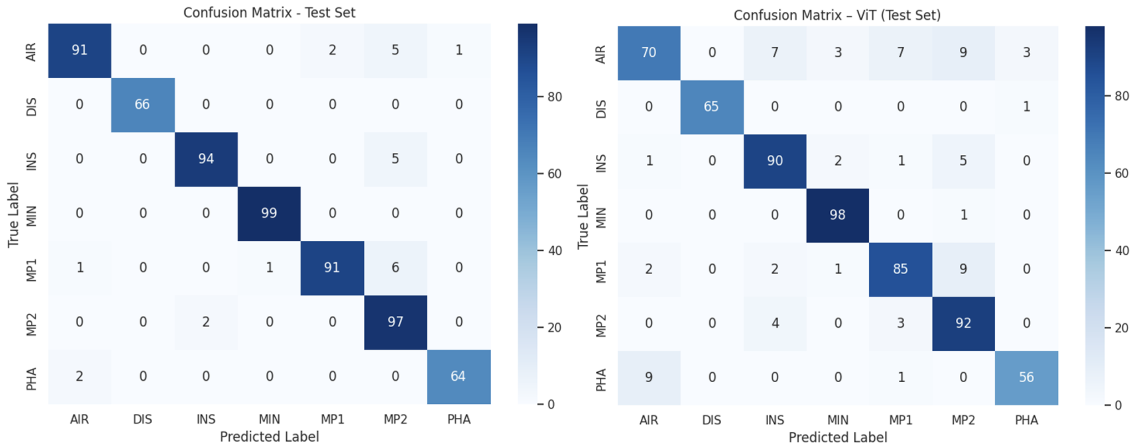

Fig. 3: Confusion Matrix - Results for VGG-16 and Transformer (Block = 30 ms)

TABLE II: Comparison of Classification Results for VGG-16 and Transformer (Block = 30 ms)

	VGG-16					Transformer			
Class	**Precision**	**Recall**	**F1-score**	**Support**	**Class**	**Precision**	**Recall**	**F1-score**	**Support**
AIR	0.9681	0.9192	0.9430	99	AIR	0.8537	0.7071	0.7735	99
DIS	1.0000	1.0000	1.0000	66	DIS	1.0000	0.9848	0.9924	66
INS	0.9792	0.9495	0.9641	99	INS	0.8738	0.9091	0.8911	99
MIN	0.9900	1.0000	0.9950	99	MIN	0.9423	0.9899	0.9655	99
MP1	0.9785	0.9192	0.9479	99	MP1	0.8763	0.8586	0.8673	99
MP2	0.8584	0.9798	0.9151	99	MP2	0.7931	0.9293	0.8558	99
PHA	0.9846	0.9697	0.9771	66	PHA	0.9333	0.8485	0.8889	66
Accuracy		0.9601		627	**Accuracy**		0.8868		627
Macro avg	0.9655	0.9625	0.9632	627	**Macro avg**	0.8961	0.8896	0.8906	627
Weighted avg	0.9627	0.9601	0.9605	627	**Weighted avg**	0.8886	0.8868	0.8854	627

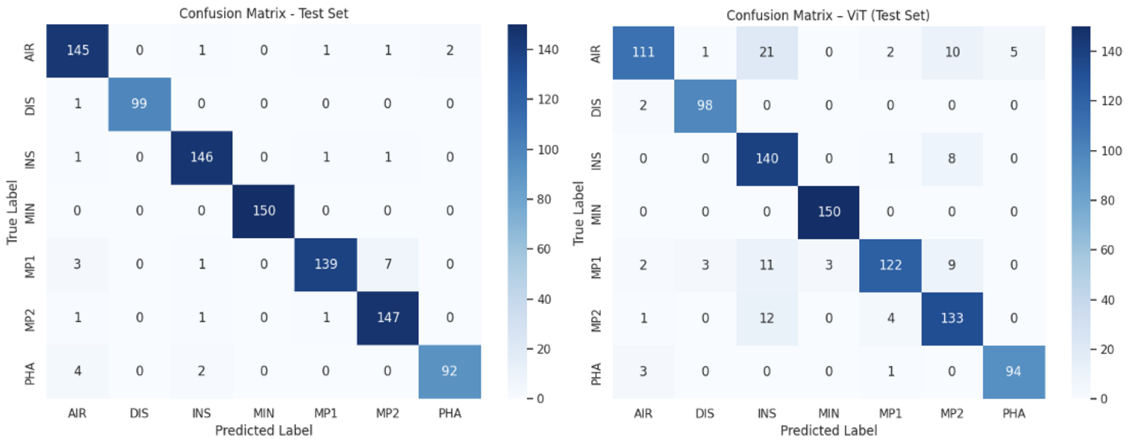

Fig. 4: Confusion Matrix - Results for VGG-16 and Transformer (Block = 20 ms)

TABLE III: Comparison of Classification Results for VGG-16 and Transformer (Block = 20 ms)

VGG-16					Transformer				
Class	**Precision**	**Recall**	**F1-score**	**Support**	**Class**	**Precision**	**Recall**	**F1-score**	**Support**
AIR	0.9355	0.9667	0.9508	150	AIR	0.9328	0.7400	0.8253	150
DIS	1.0000	0.9900	0.9950	100	DIS	0.9608	0.9800	0.9703	100
INS	0.9669	0.9799	0.9733	149	INS	0.7609	0.9396	0.8408	149
MIN	1.0000	1.0000	1.0000	150	MIN	0.9804	1.0000	0.9901	150
MP1	0.9789	0.9267	0.9521	150	MP1	0.9385	0.8133	0.8714	150
MP2	0.9423	0.9800	0.9608	150	MP2	0.8313	0.8867	0.8581	150
PHA	0.9787	0.9388	0.9583	98	PHA	0.9495	0.9592	0.9543	98
Accuracy		0.9694		947	**Accuracy**		0.8955		947
Macro avg	0.9718	0.9689	0.9700	947	**Macro avg**	0.9077	0.9027	0.9015	947
Weighted avg	0.9699	0.9694	0.9694	947	**Weighted avg**	0.9028	0.8955	0.8950	947

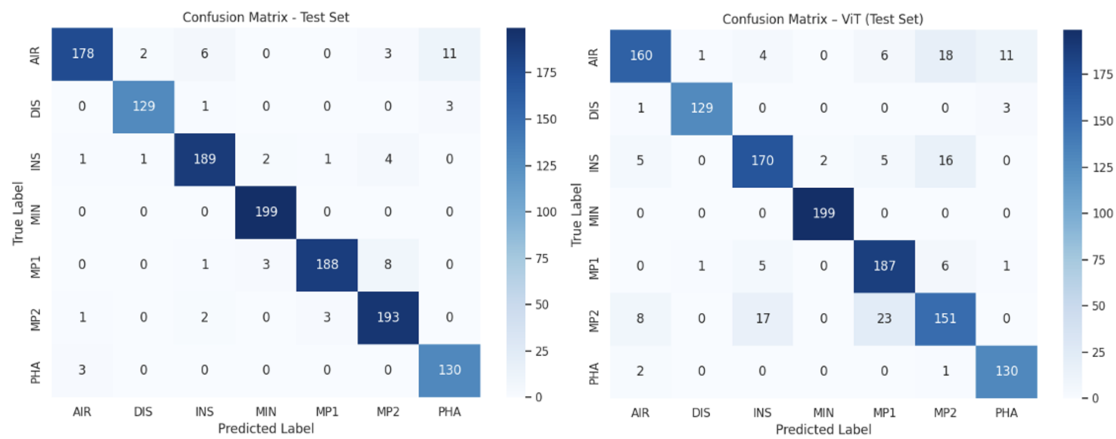

Fig. 5: Confusion Matrix - Results for VGG-16 and Transformer (Block = 15 ms)

TABLE IV: Comparison of Classification Results for VGG-16 and Transformer (Block = 15 ms)

VGG-16					Transformer				
Class	**Precision**	**Recall**	**F1-score**	**Support**	**Class**	**Precision**	**Recall**	**F1-score**	**Support**
AIR	0.9727	0.8900	0.9295	200	AIR	0.9091	0.8000	0.8511	200
DIS	0.9773	0.9699	0.9736	133	DIS	0.9847	0.9699	0.9773	133
INS	0.9497	0.9545	0.9521	198	INS	0.8673	0.8586	0.8629	198
MIN	0.9755	1.0000	0.9876	199	MIN	0.9900	1.0000	0.9950	199
MP1	0.9792	0.9400	0.9592	200	MP1	0.8462	0.9350	0.8884	200
MP2	0.9279	0.9698	0.9484	199	MP2	0.7865	0.7588	0.7724	199
PHA	0.9028	0.9774	0.9386	133	PHA	0.8966	0.9774	0.9353	133
Accuracy		0.9556		1262	**Accuracy**		0.8922		1262
Macro avg	0.9550	0.9574	0.9556	1262	**Macro avg**	0.8972	0.9000	0.8975	1262
Weighted avg	0.9566	0.9556	0.9555	1262	**Weighted avg**	0.8926	0.8922	0.8913	1262

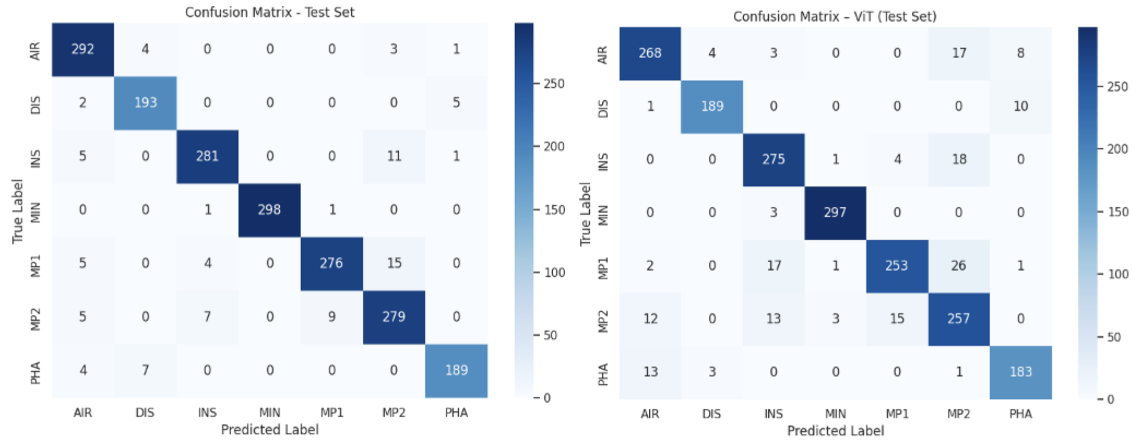

Fig. 6: Confusion Matrix - Results for VGG-16 and Transformer (Block = 10 ms)

TABLE V: Comparison of Classification Results for VGG-16 and Transformer (Block = 10 ms)

VGG-16					Transformer				
Class	**Precision**	**Recall**	**F1-score**	**Support**	**Class**	**Precision**	**Recall**	**F1-score**	**Support**
AIR	0.9329	0.9733	0.9527	300	AIR	0.9054	0.8933	0.8993	300
DIS	0.9461	0.9650	0.9554	200	DIS	0.9643	0.9450	0.9545	200
INS	0.9590	0.9430	0.9509	298	INS	0.8842	0.9228	0.9031	298
MIN	1.0000	0.9933	0.9967	300	MIN	0.9834	0.9900	0.9867	300
MP1	0.9650	0.9200	0.9420	300	MP1	0.9301	0.8433	0.8846	300
MP2	0.9058	0.9300	0.9178	300	MP2	0.8056	0.8567	0.8304	300
PHA	0.9643	0.9450	0.9545	200	PHA	0.9059	0.9150	0.9104	200
Accuracy		0.9526		1898	**Accuracy**		0.9073		1898
Macro avg	0.9533	0.9528	0.9529	1898	**Macro avg**	0.9113	0.9095	0.9099	1898
Weighted avg	0.9531	0.9526	0.9526	1898	**Weighted avg**	0.9088	0.9073	0.9075	1898

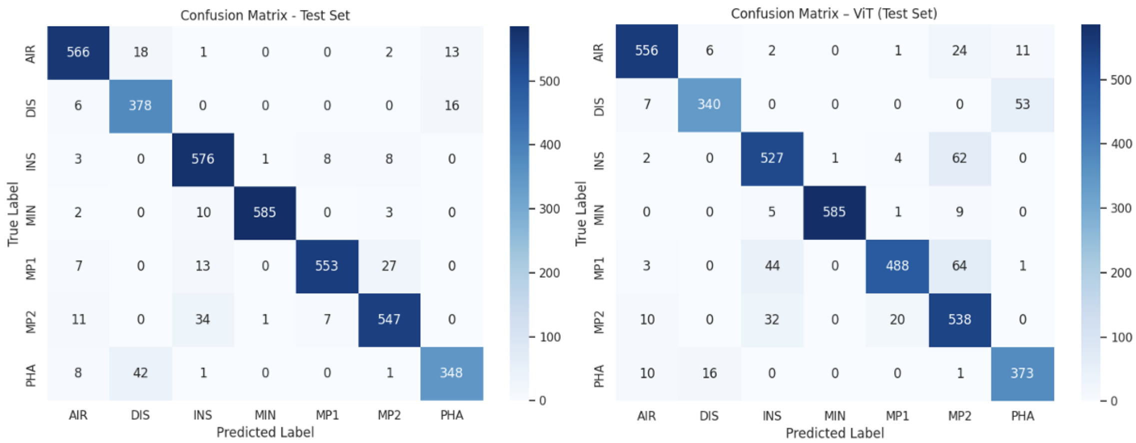

Fig. 7: Confusion Matrix - Results for VGG-16 and Transformer (Block = 5 ms)

TABLE VI: Comparison of Classification Results for VGG-16 and Transformer (Block = 5 ms)

VGG-16					Transformer				
Class	**Precision**	**Recall**	**F1-score**	**Support**	**Class**	**Precision**	**Recall**	**F1-score**	**Support**
AIR	0.9386	0.9433	0.9410	600	AIR	0.9456	0.9267	0.9360	600
DIS	0.8630	0.9450	0.9021	400	DIS	0.9392	0.8500	0.8924	400
INS	0.9071	0.9664	0.9358	596	INS	0.8639	0.8842	0.8740	596
MIN	0.9966	0.9750	0.9857	600	MIN	0.9983	0.9750	0.9865	600
MP1	0.9736	0.9217	0.9469	600	MP1	0.9494	0.8133	0.8761	600
MP2	0.9303	0.9117	0.9209	600	MP2	0.7708	0.8967	0.8290	600
PHA	0.9231	0.8700	0.8958	400	PHA	0.8516	0.9325	0.8902	400
Accuracy		0.9360		3796	**Accuracy**		0.8975		3796
Macro avg	0.9332	0.9333	0.9326	3796	**Macro avg**	0.9027	0.8969	0.8977	3796
Weighted avg	0.9374	0.9360	0.9361	3796	**Weighted avg**	0.9035	0.8975	0.8984	3796

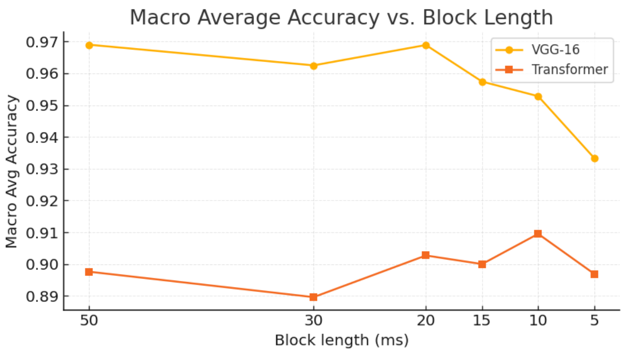

Fig. 8: Macro Accuracy per Block Length

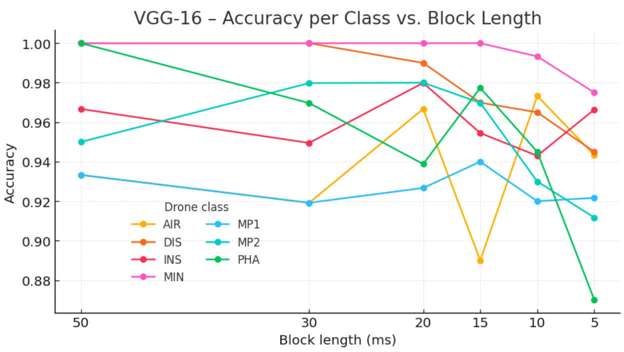

Fig. 9: Accuracy per Class - VGG-16 Model

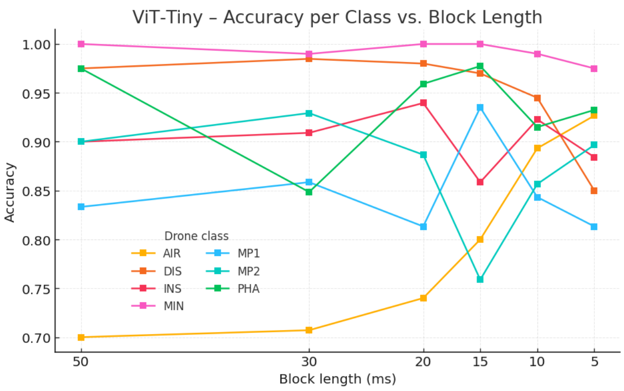

Fig. 10: Accuracy per Class - ViT-Tiny Model

Although the **VGG-16** model excelled in performance metrics, the **ViT-Tiny** model has a significantly smaller footprint, as shown in Table VII, and may be the more suitable option for embedded implementations on a Raspberry Pi. Embedded models can offer various benefits in drone classification scenarios involving autonomous drones, where weight, power consumption, and response time are critical constraints.

TABLE VII: Comparison of model efficiency metrics

Model	Params	FLOPs	Latency (ms)	Memory (GB)	Thr (img/s)	Size (MB)
VGG-16	138 M	15.5 G	12.4 ± 0.3	1.8	80	528
ViT-Tiny	11 M	1.3 G	100 ± 5	0.2	10	35

B. Comparative Analysis

The VGG-16 model with a 20 ms block achieves the best performance (approx 97 % accuracy/F1), outperforming all classical methods. The ViT-Tiny with a 10 ms block reaches about 91 %, i.e., approximately 11 pp above k-NN and 2 pp above LR in the same domain, but still 6 pp below VGG. Traditional classifiers (LR, k-NN) lag by 8–17 pp compared to the best deep models when operating on spectrograms.

TABLE VIII: Model performance

Model	Metric	Spectrogram (%)
LR [13]	Acc	88.6 ± 0.8
	F1	88.6 ± 0.8
k-NN [13]	Acc	79.7 ± 0.8
	F1	79.7 ± 0.8
VGG-16 (20 ms)	Acc	96.9
	F1	96.9
ViT-Tiny (10 ms)	Acc	90.7
	F1	90.7

VI. Conclusion

In summary, on the *DroneDetect* corpus for drone RF classification, deep models surpassed classical baselines. With 20 ms blocks, VGG-16 reached $\sim$ 97% accuracy/F1; ViT-Tiny achieved $\sim$ 91%, up to 17 percentage points above linear regression and k-NN. VGG-16 offers top precision but is heavy (138 M params; 15.5 G FLOPs), whereas ViT-Tiny is lean (11 M; 1.3 G), suiting energy/memory/latency-constrained embedded use. Thus, choosing architecture and block length entails a performance–efficiency trade-off.

References

[1] DJI, “The history of drones: From world war ii to the digital age,” May 2024, published on May 2, 2024. [Online]. Available: https://blog.lojadji.com.br/historia-dos-drones/

[2] J. Harper. (2024, March) Army seeks over us$ 400 million in fiscal year 2025 for counter-drone systems. Published on March 11, 2024. [Online]. Available: https://defensescoop.com/2024/03/11/army-counter-drone-systems-funding-fiscal-2025/

[3] “Radio frequency signal-based drone classification with frequency domain gramian angular field and convolutional neural network,” *Drones*, vol. 8, no. 9, pp. 511–511, 2024.

[4] T. Gaikwad and S. Dhavale, “Rf based enhanced drone detection and classification using short time fourier transform,” 2024.

[5] “Drone classification from rf fingerprints using deep residual nets,” 2021, pp. 548–555.

[6] “Radio frequency-based drone detection and classification using deep learning algorithms,” 2023, pp. 1–6.

[7] “Review of rf-based drone classification: Techniques, datasets, and challenges,” *VojnotehniÄki Glasnik*, vol. 72, no. 2, pp. 764–789, 2024.

[8] A. Singh, V. Sharma, and K. Rawat, “Classification of rf fingerprint signals from uav controller using machine learning techniques,” pp. 1–5, 2023.

[9] “Leveraging rf signal transformation and sigmoid calibration for optimized drone class prediction,” pp. 170–175, 2024.

[10] M. Rakesh and G. V. R. Reddy, “Combined rf-based drone detection and classification,” *International Journal for Research in Applied Science and Engineering Technology*, 2023.

[11] I. Nemer, T. Sheltami, I. Ahmad, A. U. H. Yasar, and M. A. R. Abdeen, “Rf-based uav detection and identification using hierarchical learning approach,” *Sensors*, vol. 21, p. 1947, 2021, [CrossRef] [PubMed]. [Online]. Available: https://doi.org/10.3390/s21061947

[12] C. J. Swinney and J. C. Woods, “DroneDetect Dataset: A Radio Frequency Dataset of Unmanned Aerial System (UAS) Signals for Machine Learning Applications,” https://ieee-dataport.org/open-access, 2022, submitted by: Carolyn Swinney, Last updated: Tue, 05/17/2022 - 22:21. [Online]. Available: https://ieee-dataport.org

[13] ——, “Rf detection and classification of unmanned aerial vehicles in environments with wireless interference,” in *2021 International Conference on Unmanned Aircraft Systems (ICUAS)*, 2021, pp. 1494–1498.