Yuzuncu Yil University
Journal of the Institute of Natural & Applied Sciences

https://dergipark.org.tr/en/pub/yyufbed

Research Article

Drone Classification with a Hybrid Deep Learning Approach Based on Mel-Spectrogram Representation

Nida KUMBASAR[1,3], Rabiye KILIÇ[*2,3], Emin Argun ORAL[3], Ibrahim Yucel OZBEK[3,4]

[1]TÜBİTAK BİLGEM, Informatics and Information Security Research Center, 41470, Kocaeli, Türkiye
[2]Ataturk University, Faculty of Engineering, Department of Computer Engineering, 25240, Erzurum, Türkiye
[3] Ataturk University, Faculty of Engineering, Department of Electrical and Electronics Engineering, 25240, Erzurum, Türkiye
[4]Ataturk University, Faculty of Engineering, Department of Artificial Intelligence and Data Engineering, 25240, Erzurum, Türkiye
*Corresponding author e-mail: rabiyekilic@atauni.edu.tr

Abstract: With technological advancements, the use of drones has become increasingly widespread in both civilian and military sectors in recent years. There is a need for technologies that can detect and identify the presence, type, or flight mode of a drone with remote sensing signals in situations that pose a security threat. This study is based on the classification of radio frequency (RF) signals from various drones under different flight modes using Mel-spectrogram representations. Within the scope of the study, the VGG19 deep learning model is used as both a classifier and a feature extractor for the SVM classifier. On the other hand, the study compares the performance of RF signals in low and high frequency bands separately and in concatenated versions. In the results obtained in the experimental studies, the VGG19+SVM hybrid model showed the highest performance over the Mel-spectrogram of the concatenated low and high (L+H) frequencies. The accuracy performances were 100% in the 2-Class problem where drone presence was detected (drone -no drone), 90.78% in the 4-Class problem where drone types were classification (Bebop-AR-Phantom-no drone), and 86.6% in the 10-Class problem where drone modes were obtained.

Keywords: Drone classification, Hybrid model, Mel-spectrogram, RF signal

Mel-Spektrogram Temsiline Dayalı Hibrit Derin Öğrenme Yaklaşımı ile Drone Sınıflandırması

Öz: Teknolojik gelişmelerle birlikte son yıllarda hem sivil hem de askeri sektörde dron kullanımı giderek yaygınlaşmaktadır. Güvenlik tehdidi oluşturan durumlarda uzaktan algılama sinyalleriyle bir dronun varlığını, türünü veya uçuş modunu tespit edip tanımlayabilen teknolojilere ihtiyaç duyulmaktadır. Bu çalışma, Mel-spektrogram gösterimleri kullanılarak farklı uçuş modları altında çeşitli dronlardan gelen radyo frekans (RF) sinyallerinin sınıflandırılmasına dayanmaktadır. Çalışma kapsamında, VGG19 derin öğrenme modeli hem bir sınıflandırıcı hem de SVM sınıflandırıcısı için bir özellik çıkarıcı olarak kullanılmıştır. Öte yandan çalışma, RF sinyallerinin düşük ve yüksek frekans bantlarındaki performansını ayrı ayrı ve birleştirilmiş versiyonlarda karşılaştırmaktadır. Deneysel çalışmalarda elde edilen sonuçlarda, VGG19+SVM hibrit modeli, birleştirilmiş düşük ve yüksek (L+H) frekansların Mel-spektrogramı üzerinde en yüksek performansı göstermiştir. dron varlığının tespit edildiği 2-Sınıf problemde (dron var - dron yok) doğruluk performansları %100, dron tiplerinin sınıflandırıldığı 4-Sınıf problemde (Bebop-AR-Phantom-dron yok) %90.78 ve dron modlarının elde edildiği 10-Sınıf problemde ise %86.6 olarak gerçekleşmiştir.

Anahtar Kelimeler: Drone sınıflandırma, Hibrit model, Mel-spektrogram, RF sinyal

Received: 29.12.2024
Accepted: 16.05.2025

How to cite: Kumbasar, N., Kılıç, R., Oral, E. A., & Özbek, I., Y. (2025). Drone classification with a hybrid deep learning approach based on mel-spectrogram representation. *Yuzuncu Yil University Journal of the Institute of Natural and Applied Sciences*, *30*(2), 608-620. https://doi.org/10.53433/yyufbed.1609294

1. Introduction

Unmanned aerial vehicles (UAVs), commonly known as drones, have seen a significant rise in popularity in recent years, being used extensively in various civil, commercial, and military sectors. They also pose a number of threats that can cause serious disruptions to airspace security (Taha & Shoufan, 2019). The potential for illegal airspace invasions, privacy violations, and security threats has increased the importance of examining the presence of drones in the air and determining the category of drones.

Due to their cost-efficiency, ease of operation, long-range capabilities, high mobility, and ability to carry payloads, drones have become essential tools in multiple industries.

Despite these benefits, their extensive deployment has introduced significant challenges concerning the security, privacy, and safety of operational environments. The electronic components within unmanned aerial vehicles (UAVs) emit electromagnetic energy that can be detected using radio frequency (RF) sensors. This capability allows RF-based technologies to identify and monitor UAV activity by capturing and analyzing wireless communication signals. Specifically, UAV detection can be achieved by intercepting the RF signals exchanged between the drone and its ground control station, providing a reliable method for identifying unauthorized aerial activities in sensitive areas (Seidaliyeva et al., 2023).

In drone detection and classification with machine learning (ML), pattern recognition can be performed using RFs, optical, and acoustic signals that are completely imperceptible to humans. Recent research has demonstrated the effectiveness of deep learning (DL) techniques in classifying drones using RF signals (Zhang et al., 2023).

DL is a state-of-the-art approach to achieving satisfactory results in computer vision and pattern recognition. Drone detection and classification based on visual data with convolutional neural networks (CNN)s is still in its infancy. The requirement for extensive labeled data to develop robust DL models, combined with the limited availability of public datasets, hampers progress in this field. To address this challenge, researchers have turned to transfer learning.

In recent years, an increasing amount of research has been focused on UAV detection, tracking, and classification, utilizing various sensor types, including radar sensors (Yousaf et al., 2022; Narayanan et al., 2023; Yan et al., 2023a), RF sensors (Kılıç et al., 2022; Kumbasar et al., 2022; Mandal & Satija, 2023), audio sensors (Anwar et al., 2019; Al-Emadi et al., 2021; Utebayeva et al., 2023), and camera sensors (Singha & Aydin, 2021; Zhao et al., 2022; Aydin & Singha, 2023). In addition, bimodal and multimodal sensor fusions are also an emerging methodology in this field (Seidaliyeva et al., 2023).

Although recent studies have explored the use of DL and time-frequency representations for RF-based drone classification, most approaches either focus on a limited number of drone types or do not combine low and high frequency information effectively. This study addresses this gap by proposing a hybrid architecture that enhances classification performance across different frequency bands and problem complexities (2-Class, 4-Class, and 10-Class), thus contributing to more robust and accurate drone detection and classification systems.

1.1. Related work

The widespread adoption of high-performance computing units and the rapid growth of data have led to the increasing popularity of DL-based methods, reducing reliance on the quality of manually crafted features. Leveraging the success of DL techniques in extracting features from 2D data, the transformation of signals from the time domain to the time-frequency domain has gained traction (Ezuma et al., 2019; Alam et al., 2023). This approach aims to minimize the impact of noise outside the signal's frequency band, thereby enhancing the model's ability to identify patterns accurately.

In recent years, with the rapid increase in drone usage across different sectors, significant attention has been paid to developing techniques for detecting and classifying drones. Ezuma et al. (2019) explored 15 statistical features, including mean, standard deviation, and entropy, and used neighbor component analysis (NCA) to reduce the dimensionality of these features. The reduced feature set was then analyzed using three machine learning (ML) algorithms: discriminant analysis (DA), support vector machine (SVM), and neural network (NN). Alam et al. (2023) proposed a model for signal detection and classification by processing time-domain RF signals using a 1D-based SqueezeNet architecture. Basak et al. (2021) applied the short-time Fourier transform (STFT) to convert RF drone signals into time-frequency (TF) representations, which were then used in a ResNet model to evaluate detection performance under varying signal-to-noise ratios (SNRs). Medaiyese et al. (2022) utilized wavelet transformations to represent drone signals in the time-frequency domain and compared the performance of discrete wavelet transform (DWT), continuous wavelet transform (CWT), and wavelet scattering transform (WST). They concluded that WST combined with SqueezeNet yielded the best signal identification results. Yan et al. (2023b) introduced a hybrid transformer model that integrates a CNN-based architecture for generating time-frequency features and incorporates self-attention mechanisms to improve the classification of drone RF signals. Mohammed et al. (2023) converted drone RF signals into Mel-spectrograms and utilized the YAMNet neural network for effective drone detection and classification. The requirement for extensive labeled data to develop robust DL models, combined with the limited availability of public datasets, hampers progress in this field. To address this challenge, researchers have turned to transfer learning. Compared to camera images, Mel-spectrograms allow detection in dark environments regardless of visual conditions. Due to disadvantages such as background noise affecting the spectrogram representation, gathering information about drones from Mel-spectrograms alone may be insufficient.

For this purpose, this study aims to extract features from Mel-spectrogram images with the help of CNN architectures and classify them with SVM. The general framework of the study consists of combining low and high frequencies of drone RF signals in the time domain, creating segment-based Mel-spectrograms, using the VGG19 robust classifier network as a feature extractor and finally feeding it to SVM. In each of these steps, it is aimed to increase the drone classification performance by valorizing the information.

Although DL techniques have been used in RF-based drone classification in recent literature, most approaches rely solely on end-to-end learning and often focus on a limited number of classes or signal types. In contrast, this study introduces a hybrid classification framework that combines the strength of pre-trained CNN architectures for feature extraction with the generalization ability of classical machine learning algorithms. The approach is further enhanced by the use of fused low and high-frequency RF signals in the form of Mel-spectrograms.

The main contributions of this study can be summarized as follows:

- A hybrid DL model is proposed that integrates VGG19 for feature extraction with SVM for classification of drone RF signals.
- The study utilizes Mel-spectrogram representations of both low and high frequency RF signals, including their concatenated (L+H) forms, to enhance the robustness of classification.
- A comprehensive evaluation is conducted using the DroneRF dataset, addressing binary (2-Class), multi-class (4-Class), and fine-grained (10-Class) drone classification problems.
- The proposed method achieves state-of-the-art performance, particularly in the 10-Class problem, demonstrating significant improvement over existing approaches in the literature.

The structure of the paper is as follows: Section 2 outlines the dataset utilized in the study, the data preprocessing steps, and the development of the model. Section 3 presents the experimental setting, performance metrics, and experimental results, as well as a comparison between the models. In Section 4, a comparison with the literature on the same dataset is used, and the importance of the study is emphasized. Finally, conclusions and future work are presented in Section 5.

2. Material and Methods

This section describes the materials and methods used in drone detection and classification.

2.1. DroneRF database

In this study, the DroneRF dataset introduced to the literature by Allahham et al. (2019) is used. The DroneRF dataset consists of a total of 454 records containing AR, Bebop, and Phantom drone signals at low and high frequencies, as well as background information. The dataset is about 40 GB in size and consists of 10.25 seconds of RF drone recordings for each flight mode and about 5.25 seconds for RF background activities.

Table 1. Definitions of drone modes

Definition	Mode
Drone mode on	Mode 1
Drone hovering	Mode 2
Drone recording video	Mode 3
Drone not recording	Mode 4

The DroneRF dataset was created by recording the drones in different modes, as shown in Table 1. The Bebop and AR drones were recorded in four different modes: open, hovering, video recording, and flying, while the Phantom drone was recorded in only one mode. Table 2 shows the number of segments in the low and high frequency bands from each drone in different modes.

Table 2. Number of dataset records

Dataset Records	Low	High
B. mode 1	21	21
B. mode 2	21	21
B. mode 3	21	21
B. mode 4	21	21
AR mode 1	21	21
AR mode 2	21	21
AR mode 3	21	21
AR mode 4	18	18
P. mode 1	21	21
No Drone (Background)	41	41
Total	**227**	**227**

Figure 1 presents a visual description of the 2-Class, 4-Class, and 10-Class problems considered in this study.

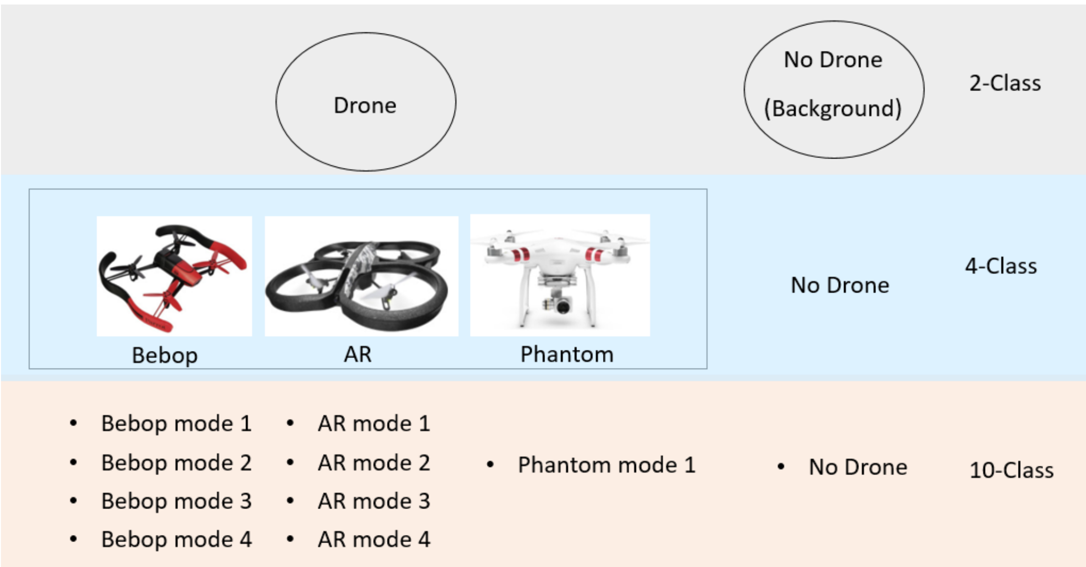

Figure 1. Details of DroneRF dataset.

The RF signals at each frequency in the data set were divided into 100 segments with 75% overlap to increase the number of data and to solve the classification problem based on shorter duration signal information. On the other hand, the low and high signals are concatenated as in Equation 1 to utilize the information of both signals at the same time.

$$\mathbf{z}_l=[x_1,x_2,x_3...x_n]$$

$$\mathbf{z}_h=[y_1,y_2,y_3....y_n] \tag{1}$$

$$\mathbf{z}_l=\text{concat}[\mathbf{z}_l,\mathbf{z}_h]=[x_1,x_2,x_3...x_n,y_1,y_2,y_3....y_n]$$

where is the representation of the low signal and $\mathbf{z}_h$ is the representation of the high signal in time.

2.2. Mel-Spectrogram

In audio signal analysis, spectrograms offer valuable insight into the distribution of frequency energy over time (Allahham et al., 2019). While general spectrograms capture extensive information about the frequency content of the data, Mel-spectrograms are specifically based on the Mel scale, which aligns with human auditory perception. As a result, Mel-spectrograms provide an effective representation of frequencies that are most relevant to human hearing.

The stability of Mel-spectrograms despite variations in audio signals underscores their effectiveness as a feature extraction method for audio data (Alam et al., 2023). By applying the Mel scale transformation, acoustic features are captured more efficiently, allowing for a better focus on the most meaningful aspects of the signal. The Mel scale frequency is mathematically calculated as shown in Equation 2.

$$f_{mel} = 2595log_{10}\left(1+\frac{f_{Hz}}{700}\right) \tag{2}$$

where f_{Hz} is the frequency of the received signal and f_{mel} is the corresponding frequency in Mel scale. The Mel-spectrum involves the application of a Mel filter bank to the frequency domain of the signal. Each filter in the Mel filter bank consists of triangular window functions H_k centered at a specific point in the frequency domain (Equation 3). These filters divide the signal into different frequency bands.

$$H_k(m) = \begin{cases} 0 & for\ m < f(k-1) \\ \frac{m-f(k-1)}{f(k)-f(k-1)} & for\ f(k-1) \leq m \leq f(k) \\ \frac{f(k+1)-m}{f(k+1)-f(k)} & for\ f(k) < m \leq f(k+1) \\ 0 & for\ m > f(k+1) \end{cases} \tag{3}$$

where $H_k(m)$ is the triangular window function, m is the number of filters and k ranges from 0 to $(m-1)$.

The Mel-spectrum is then obtained by squaring the discrete Fourier transform (DFT) coefficients as shown in Equation 4 and applying the corresponding triangular window functions from the Mel filter set.

$$X_{Mel}(x) = \sum_{m=0}^{N-1}[|X(m)|^2 H_k(m)] \tag{4}$$

Figure 2 shows the Mel-spectrogram representations of the RF signals of the DroneRF dataset in the low and high frequency bands.

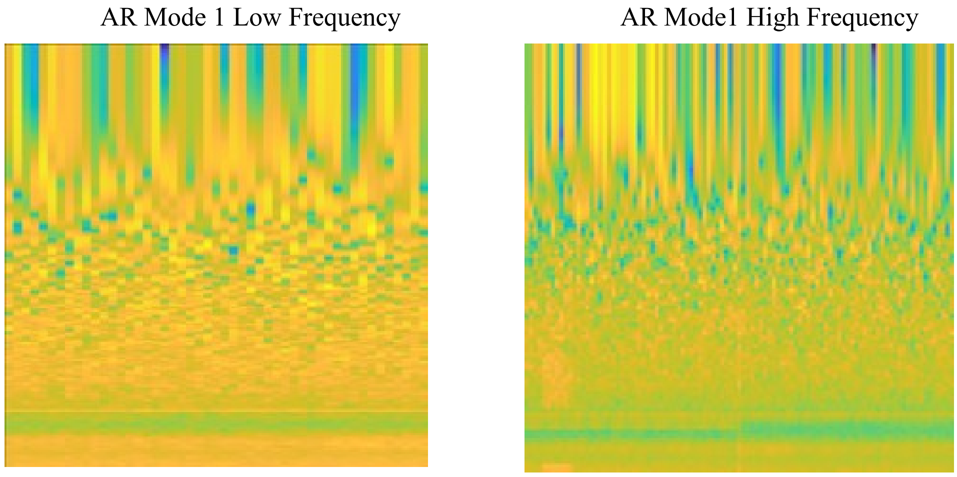

Figure 2. Mel-spectrogram equivalent of an example AR mode 1 RF signal.

2.3. VGG19+SVM hybrid model

This study proposes a hybrid DL model that combines the VGG19 network's feature extraction capabilities with the classification power of SVM to analyze drone RF Mel-spectrograms. The model is designed to enhance both the accuracy and robustness of drone detection and classification tasks by integrating these complementary techniques.

The VGG19 network (Simonyan & Zisserman, 2014), a deep CNN developed by the Visual Geometry Group, garnered widespread recognition for its outstanding performance in the 2014 ImageNet Large Scale Visual Recognition Competition (ILSVRC). Comprising 19 layers, VGG19 includes convolutional, max-pooling, and fully connected layers. Its convolutional and fully connected layers are equipped with trainable parameters, enabling adaptability during training, while its max-pooling layers reduce input data dimensions by retaining only the most critical features. Due to its exceptional performance in tasks such as image classification, object detection, and segmentation, VGG19 has become a widely adopted model for transfer learning, making it highly effective in extracting relevant features from drone RF signals.

SVM, introduced by (Cortes & Vapnik, 1995), offer a robust approach to classification by determining an optimal hyperplane that separates data into distinct categories. The placement of this hyperplane maximizes the margin between data points closest to it, known as support vectors. These vectors play a crucial role in defining the hyperplane's orientation and position, ensuring the classifier achieves high accuracy. By focusing on maximizing the margin, SVM generalizes effectively to unseen data, making it particularly useful for high-dimensional datasets such as drone RF signals.

The proposed hybrid model, shown in Figure 3, processes RF signals by generating their Mel-spectrogram representations, which capture the time-frequency characteristics of the signals. These spectrograms, derived from various frequency bands, including low-frequency, high-frequency, and a combination of both, are fed into the VGG19 network. Within VGG19, convolutional and fully connected layers extract critical features from the input data. These feature vectors are subsequently passed to the SVM classifier, which performs the final classification, differentiating between various drone types based on their RF Mel-spectrogram characteristics. By integrating VGG19's advanced feature extraction capabilities with SVM's robust classification approach, this hybrid model provides an effective solution for analyzing complex RF signals. The approach not only improves classification accuracy but also demonstrates substantial potential for practical applications, particularly in security and surveillance domains where reliable drone detection is essential. This combination of techniques underscores the potential of hybrid DL models in addressing challenges associated with drone RF signal analysis while maintaining accuracy and efficiency.

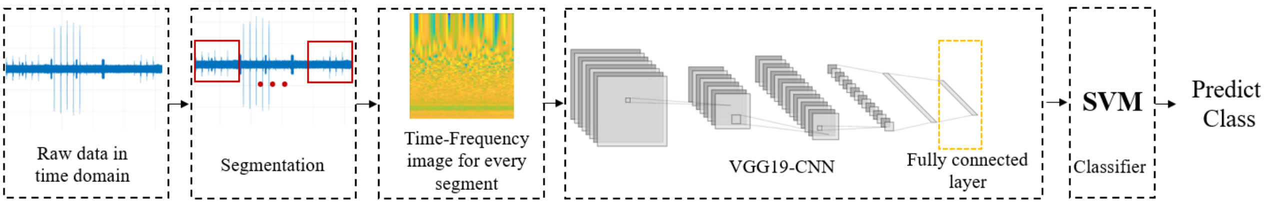

Figure 3. The proposed model diagram.

In the Mel-spectrogram generation phase, each RF segment was processed using a Hamming window of 1024 samples with a hop size of 512 and 128 Mel bands. The resulting time-frequency images were resized to fit the input dimensions of the pre-trained CNN models. For training, we used a batch size of 32, learning rate of 0.0001, and 20 epochs. The CNN's last fully connected layer (prior to classification) was used to extract 4096-dimensional feature vectors, which were then input to the SVM classifier with an RBF kernel.

3. Results

In this section, evaluation criteria of experimental studies, experimental setup, and experimental results are explained.

3.1. Evaluation criteria

In this study, accuracy is used to evaluate the performance of the proposed methods by measuring the proportion of correct predictions, both positive and negative, made by the model. It provides an overall assessment of the model's classification capability.

Confusion matrix (Figure 4) provides four key components to evaluate the model's prediction performance: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). True Positive represents cases where the model correctly predicts a positive outcome, such as accurately detecting the presence of a drone. True Negative refers to instances where the model correctly identifies a negative case, such as determining that no drone is present. On the other hand, False Positive occurs when the model incorrectly predicts a positive outcome, mistaking a non-drone situation for a drone presence. False Negative represents cases where the model fails to identify an existing drone, incorrectly predicting it as absent. These components are crucial for calculating performance metrics like accuracy, precision, and recall, offering a comprehensive understanding of the model's strengths and weaknesses. Equations 5, 6, 7 and 8 present the definition of Accuracy, Precision Recall and F1 score metrics with TP, TF, FN, FP respectively (Kılıç et al., 2022).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{5}$$

$$Precision = \frac{TP}{TP+FP} \tag{6}$$

$$Recall = \frac{TP}{TP+FP} \tag{7}$$

$$F1\ score = 2\left(\frac{precision\ x\ recall}{precision + recall}\right) \tag{8}$$

		Positive	**Negative**
Output Class	**Positive**	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)
		Target Class	

Figure 4. Confusion matrix of a classification model.

3.2. Experimental setup

The experimental studies were conducted using an i9-7900X central processing unit paired with two GeForce RTX 2080Ti GPUs to optimize the execution of matrix multiplication and convolution operations. All algorithms were developed and executed on the MATLAB® 2019b platform, running on the Ubuntu 18.04.4-LTS Linux operating system.

3.3. Experimental results

In this section, drone classification results with 2-4-10 classes are presented in two separate tables.

Table 3 shows the VGG19 classification results based on Mel-spectrogram images in low and high frequency bands. When the results in Table 3 are analyzed, the highest results for 2-Class are 99.77%, 99.71%, and 98.44% in High (H), Low+High (L+H), and Low (L) frequency bands, respectively, and the highest results for 4 Class are 87.51%, 77.51%, and 74.64% in (L+H), High (H), and Low (L) frequency bands, respectively. For 10 Class, and 84.45%, 76.94%, and 71.53% for the Low+High (L+H), High (H), and Low (L) frequency bands, respectively.

Table 3. VGG19 experimental results (accuracy %)

Classification Mode	L band	H band	L+H band
2-Class	98.44	**99.77**	99.71
4-Class	74.64	77.51	**87.51**
10-Class	71.53	76.94	**84.45**

Upon analyzing the results of the VGG19+SVM hybrid model in Table 4, it is evident that the performance is enhanced when Mel-spectrograms are input into the VGG19 network, and the features extracted from the fully connected layer are classified using the SVM. In the 2-Class classification, the accuracy improvement with the L band is 0.10%, with the H band is 0.02%, and with the L+H band is 0.29%. For the 4-Class classification, the improvement with the L band is 3.00%, with the H band is 2.04%, and with the L+H band is 3.70%. In the 10-Class classification, the L band shows an improvement of 0.88%, the H band shows a slight improvement of 0.03%, and the L+H band provides a notable improvement of 2.55%. These results highlight the varying degrees of improvement across different bands and classification tasks, with the L+H band generally offering the highest accuracy gains.

Table 4. VGG19+SVM hybrid model experimental results (accuracy %)

Classification Mode	L band	H band	L+H band
2-Class	98.54	99.79	**100**
4-Class	76.87	79.09	**90.78**
10-Class	72.16	76.96	**86.6**

Table 5 presents the weighted by the number of data average Precision Recall and F1 score values of the VGG19+SVM hybrid model for concatenated L+H.

Table 5. VGG19+SVM hybrid model experimental results for L+H band (%)

Classification Mode	Accuracy	Precision	Recall	F1 score
2-Class	100	100	100	100
4-Class	90.78	81.20	75.40	73.60
10-Class	86.60	78.70	79.60	82.00

Figures 5-6-7 show the performance comparisons of VGG19 and VGG19+SVM models according to the Accuracy metric for 2-Class, 4-Class, 10-Class respectively.

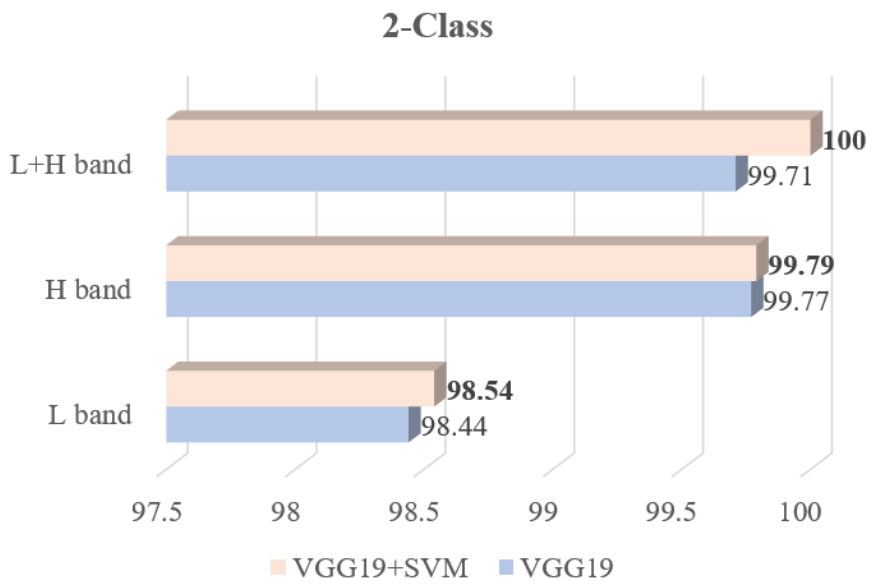

Figure 5. Drone detection classification performance.

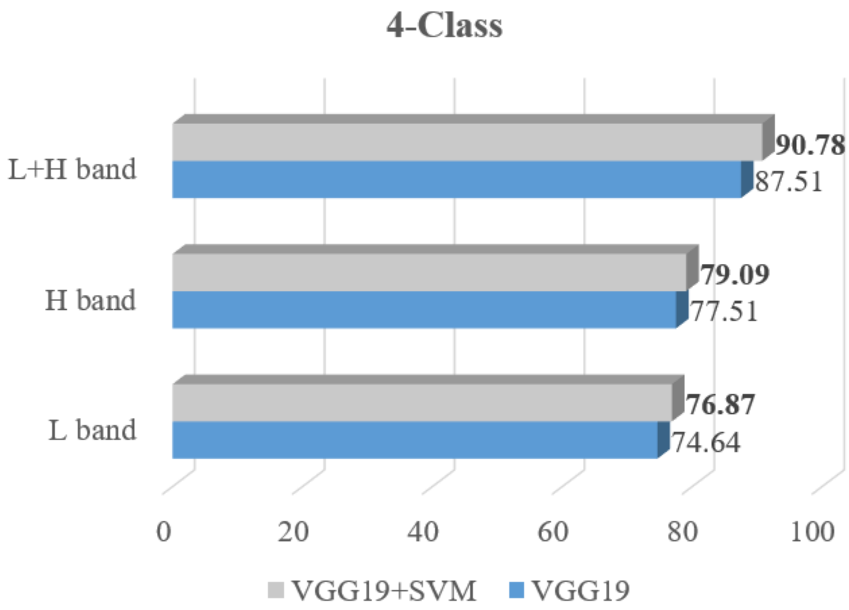

Figure 6. Drone type classification performance.

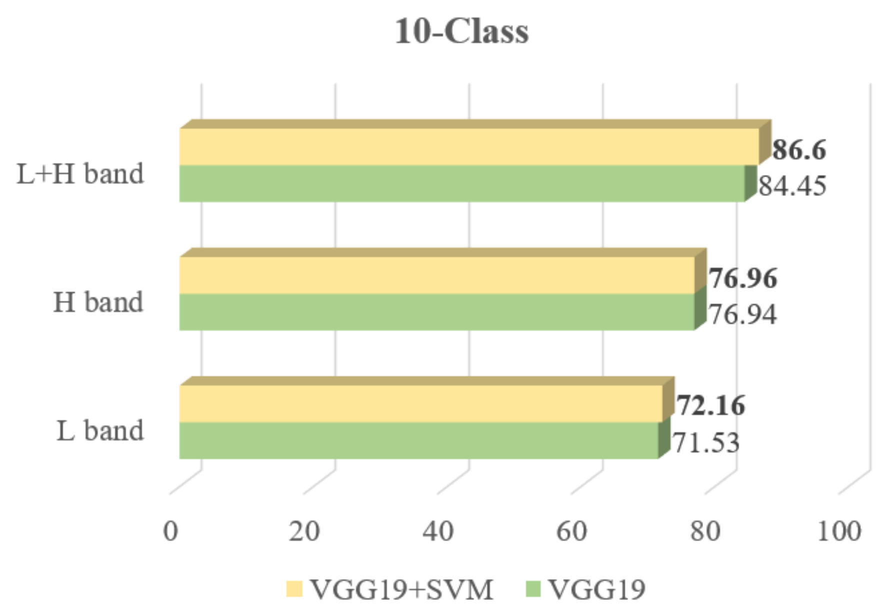

Figure 7. Drone mode classification performance.

This study addresses key limitations in the existing literature by expanding the classification to multiple drone types (AR, Bebop, Phantom) and their respective flight modes. The proposed model is evaluated in three classification settings: (i) drone presence detection (2-Class), (ii) drone type classification (4-Class), and (iii) drone type and mode recognition (10-Class). The experimental results show that the hybrid model achieved 100% accuracy for drone detection, 90.78% for drone type classification, and 86.6% for drone type+mode classification. These results demonstrate the robustness of our approach across both hardware (drone variety) and operational diversity (flight modes).

4. Discussion

The increasing prevalence of unmanned aerial vehicles (UAVs) has accelerated the development of RF-based detection and classification systems. Most UAVs operate in the 2.4 GHz ISM band, emitting characteristic radio frequency (RF) signals that can be passively intercepted and analyzed to determine drone presence, type, or operational mode (Taha & Shoufan, 2019). In this study, a hybrid classification approach was proposed, combining Mel-spectrogram representations of RF signals with DL-based feature extraction and SVM-based classification. To evaluate the effectiveness of the proposed method, comprehensive experiments were conducted on the publicly available DroneRF dataset, which includes RF recordings from multiple drones (AR, Bebop, Phantom) under various flight modes. The classification tasks were structured in three levels of complexity: binary classification (drone vs. no drone), multi-class classification (drone types), and fine-grained classification (drone types with flight modes), corresponding to 2-Class, 4-Class, and 10-Class problems, respectively.

Table 6 presents a comparison of classification accuracy achieved by the proposed method and several state-of-the-art approaches from the literature. Al-Sa'd et al. (2019) employed a deep neural network (DNN) based on power spectrum features derived from DFT magnitudes, achieving 99.7%, 84.5%, and 46.8% accuracy for 2-, 4-, and 10-Class tasks, respectively. Al-Emadi and Al-Senaid (2020) developed a 1D-CNN tailored to raw signal input, slightly improving the 2-Class and 4-Class performances and significantly outperforming previous models in the 10-Class task with 59.2% accuracy. Medaiyese et al. (2021) applied XGBoost models on wavelet-transformed RF data, achieving 99.96% (2-Class), 90.73% (4-Class), and 70.09% (10-Class) accuracy, which until now represented the best performance on the dataset for the most complex task.

Table 6. Comparative accuracy (%) of the proposed method and related studies on DroneRF dataset

Literature	2-Class	4-Class	10-Class
(Al-Sa'd et al., 2019)	99.70	84.50	46.80
(Al-Emadi & Al-Senaid, 2020)	99.80	85.80	59.20
(Medaiyese et al., 2021)	99.96	90.73	70.09
The proposed method	**100**	**90.78**	**86.60**

In contrast, the proposed method introduces several improvements over prior work. First, the concatenation of low- and high-frequency RF signals captures a broader spectrum of temporal and spectral patterns. Second, the use of Mel-spectrograms transforms RF signals into image-like representations, which are more compatible with convolutional architectures and robust to noise. Finally, the hybrid architecture combines deep CNN-based feature extraction with SVM, leveraging the strengths of both methods. The experimental results demonstrate that the proposed approach achieves 100% accuracy in the 2-Class task, 90.78% in the 4-Class task, and 86.6% in the 10-Class task—surpassing all previously reported results on the same dataset, particularly with a 16.5 percentage point improvement in the 10-Class problem over the closest competing method.

These findings suggest that the hybrid use of time-frequency representations (via Mel-spectrograms) and classical machine learning classifiers, in combination with pre-trained DL architectures, can provide a highly effective solution for drone detection and classification. The consistent improvement across all task complexities confirms the generalizability and robustness of the proposed method.

5. Conclusion

The increasing prevalence of drones has highlighted the need for improved security measures, making drone detection and classification more critical than ever. Various techniques, including optical, radar, thermal, acoustic, and RF signal processing, are employed for effective detection. This study uses Mel-spectrogram representations of RF signals in both the low and high-frequency bands from the DroneRF dataset, processed separately and in combination.

For classification tasks involving drone presence (2-Class), drone type (4-Class), and both drone type and mode (10-Class), Mel-spectrograms of RF signals from both frequency bands are input into the VGG19 network. The features extracted by the VGG19 network are then classified using SVM. The experimental results show that the proposed data representation and hybrid model approach achieve high classification accuracy. Specifically, for the 2-Class classification, the model achieved an accuracy of 100%, for the 4-Class classification, the accuracy was 90.78%, and for the 10-Class classification, it reached 86.60%. These results suggest that the proposed method is promising for practical, real-world drone detection applications.

Future research will focus on developing transformer-based multimodal models, incorporating multiple input sources and diverse drone data types. Additionally, utilizing various RF drone datasets for both training and testing is expected to yield more generalized and robust results, enhancing the model's performance across different scenarios.

References

Alam, S. S., Chakma, A., Rahman, M. H., Bin Mofidul, R., Alam, M. M., Utama, I. B. K. Y., & Jang, Y. M. (2023). RF-enabled deep-learning-assisted drone detection and identification: An end-to-end approach. *Sensors*, *23*(9), 4202. https://doi.org/10.3390/s23094202

Al-Emadi, S., Al-Ali, A., & Al-Ali, A. (2021). Audio-based drone detection and identification using deep learning techniques with dataset enhancement through generative adversarial networks. *Sensors*, *21*(15), 4953. https://doi.org/10.3390/s21154953

Al-Emadi, S., & Al-Senaid, F. (2020). Drone detection approach based on radio-frequency using convolutional neural network. *2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT)*, 29–34. https://doi.org/10.1109/iciot48696.2020.9089489

Allahham, M. S., Al-Sa'd, M. F., Al-Ali, A., Mohamed, A., Khattab, T., & Erbad, A. (2019). DroneRF dataset: A dataset of drones for RF-based detection, classification and identification. *Data in Brief*, *26*, 104313. https://doi.org/10.1016/j.dib.2019.104313

Al-Sa'd, M. F., Al-Ali, A., Mohamed, A., Khattab, T., & Erbad, A. (2019). RF-based drone detection and identification using deep learning approaches: An initiative towards a large open source drone database. *Future Generation Computer Systems*, *100*, 86–97. https://doi.org/10.1016/j.future.2019.05.007

Anwar, M. Z., Kaleem, Z., & Jamalipour, A. (2019). Machine learning inspired sound-based amateur drone detection for public safety applications. *IEEE Transactions on Vehicular Technology*, *68*(3), 2526–2534. https://doi.org/10.1109/tvt.2019.2893615

Aydin, B., & Singha, S. (2023). Drone detection using YOLOv5. *Engineering*, *4*(1), 416–433. https://doi.org/10.3390/eng4010025

Basak, S., Rajendran, S., Pollin, S., & Scheers, B. (2021). Drone classification from RF fingerprints using deep residual nets. *2021 International Conference on COMmunication Systems & NETworkS (COMSNETS)*, 548–555. https://doi.org/10.1109/comsnets51098.2021.9352891

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*, 273–297. https://doi.org/10.1007/BF00994018

Ezuma, M., Erden, F., Anjinappa, C. K., Ozdemir, O., & Guvenc, I. (2019). Detection and classification of UAVs using RF fingerprints in the presence of Wi-Fi and Bluetooth interference. *IEEE Open Journal of the Communications Society*, *1*, 60–76. https://doi.org/10.1109/ojcoms.2019.2955889

Kılıç, R., Kumbasar, N., Oral, E. A., & Ozbek, I. Y. (2022). Drone classification using RF signal based spectral features. *Engineering Science and Technology, an International Journal*, *28*, 101028. https://doi.org/10.1016/j.jestch.2021.06.008

Kumbasar, N., Kılıç, R., Oral, E. A., & Ozbek, I. Y. (2022). Comparison of spectrogram, persistence spectrum and percentile spectrum based image representation performances in drone detection and classification using novel HMFFNet: Hybrid model with feature fusion network. *Expert Systems with Applications*, *206*, 117654. https://doi.org/10.1016/j.eswa.2022.117654

Mandal, S., & Satija, U. (2023). Time–frequency multiscale convolutional neural network for RF-based drone detection and identification. *IEEE Sensors Letters*, *7*(7), 1–4. https://doi.org/10.1109/lsens.2023.3289145

Medaiyese, O. O., Ezuma, M., Lauf, A. P., & Guvenc, I. (2022). Wavelet transform analytics for RF-based UAV detection and identification system using machine learning. *Pervasive and Mobile Computing*, *82*, 101569. https://doi.org/10.1016/j.pmcj.2022.101569

Medaiyese, O. O., Syed, A., & Lauf, A. P. (2021). Machine learning framework for RF-based drone detection and identification system. *2021 2nd International Conference On Smart Cities, Automation & Intelligent Computing Systems (ICON-SONICS)*, 58–64. https://doi.org/10.1109/icon-sonics53103.2021.9617168

Mohammed, K. K., Abd El-Latif, E. I., El-Sayad, N. E., Darwish, A., & Hassanien, A. E. (2023). Radio frequency fingerprint-based drone identification and classification using Mel spectrograms and pre-trained YAMNet neural. *Internet of Things*, *23*, 100879. https://doi.org/10.1016/j.iot.2023.100879

Narayanan, R. M., Tsang, B., & Bharadwaj, R. (2023). Classification and discrimination of birds and small drones using radar micro-doppler spectrogram images. *Signals*, *4*(2), 337–358. https://doi.org/10.3390/signals4020018

Seidaliyeva, U., Ilipbayeva, L., Taissariyeva, K., Smailov, N., & Matson, E. T. (2023). Advances and challenges in drone detection and classification techniques: A state-of-the-art review. *Sensors*, *24*(1), 125. https://doi.org/10.3390/s24010125

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint* arXiv:1409.1556

Singha, S., & Aydin, B. (2021). Automated drone detection using YOLOv4. *Drones*, *5*(3), 95. https://doi.org/10.3390/drones5030095

Taha, B., & Shoufan, A. (2019). Machine learning-based drone detection and classification: State-of-the-art in research. *IEEE Access*, *7*, 138669–138682. https://doi.org/10.1109/access.2019.2942944

Utebayeva, D., Ilipbayeva, L., & Matson, E. T. (2023). Practical study of recurrent neural networks for efficient real-time drone sound detection: A review. *Drones*, *7*(1), 26. https://doi.org/10.3390/drones7010026

Yan, J., Hu, H., Gong, J., Kong, D., & Li, D. (2023a). Exploring radar micro-doppler signatures for recognition of drone types. *Drones*, *7*(4), 280. https://doi.org/10.3390/drones7040280

Yan, X., Han, B., Su, Z., & Hao, J. (2023b). SignalFormer: Hybrid transformer for automatic drone ıdentification based on drone RF signals. *Sensors*, *23*(22), 9098. https://doi.org/10.3390/s23229098

Yousaf, J., Zia, H., Alhalabi, M., Yaghi, M., Basmaji, T., Shehhi, E. A., Gad, A., Alkhedher, M., & Ghazal, M. (2022). Drone and controller detection and localization: Trends and challenges. *Applied Sciences*, *12*(24), 12612. https://doi.org/10.3390/app122412612

Zhang, H., Li, T., Li, Y., Li, J., Dobre, O. A., & Wen, Z. (2023). RF-based drone classification under complex electromagnetic environments using deep learning. *IEEE Sensors Journal*, *23*(6), 6099–6108. https://doi.org/10.1109/jsen.2023.3242985

Zhao, J., Zhang, J., Li, D., & Wang, D. (2022). Vision-based anti-UAV detection and tracking. *IEEE Transactions on Intelligent Transportation Systems*, *23*(12), 25323–25334. https://doi.org/10.1109/tits.2022.3177627