

Classification and Detection of Multiple UAVs using Rational Gaussian Wavelet Neural Networks

Received: 19 Apr 2026

Accepted: 17 Aug 2026

Published online: 08 September 2026

Gergő Ungvári, Ferenc Braun, Attila Ámon, Péter Kackstädter, János Volk, Péter Kovács & Tamás Dózsa

Cite this article as: Ungvári, G., Braun, F., Ámon, A. *et al.* Classification and Detection of Multiple UAVs using Rational Gaussian Wavelet Neural Networks. *J Intell Robot Syst* (2026).
<https://doi.org/10.1007/s10846-026-02453-2>

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

Classification and detection of multiple UAVs using rational Gaussian wavelet neural networks

Gergő Ungvári^{1,2}, Ferenc Braun⁶, Attila Ámon^{1,5},
Péter Kackstädter³, János Volk⁶, Péter Kovács^{1*}, Tamás Dózsa^{2,4}

^{1*}Department of Numerical Analysis, Eötvös Loránd University,
Pázmány Péter sétány, Budapest, 1117, Hungary.

²System and Control Laboratory, HUN-REN Institute for Computer
Science and Control, Kende utca, Budapest, 1111, Hungary.

³Szent György Hang- és Filmművészeti Technikum, Lenhossék utca,
Budapest, 1096, Hungary.

⁴Department of Mathematics and Computer Science,
UniDistanceSuisse, Schinerstrasse, Brig-Glis, 3900, Switzerland.

⁵Siemens Mobility Kft., Gábor Dénes utca, Budapest, 1117, Hungary.

⁶HUN-REN Centre for Energy Research, Konkoly-Thege Miklós út,
Budapest, 1121, Hungary.

*Corresponding author(s). E-mail(s): kovika@inf.elte.hu;

Contributing authors: aele02@inf.elte.hu; braun.ferenc@ek.hun-ren.hu;
ze3vjn@inf.elte.hu; kackstadter.peter@broadcast.hu;
volk.janos@ek.hun-ren.hu; tamas.dozsa@unidistance.ch;

Abstract

The detection of unmanned aerial vehicles (UAVs) is important for the protection of civilian and military infrastructure. In this paper we propose a cost effective UAV detection system using sound signals obtained from microphones. The recorded signals are passed through a signal processing pipeline which employs interpretable adaptive feature extractors using so-called rational Gaussian wavelets. These adaptive wavelet transformations are embedded into and trained together with an underlying small neural network which detects and classifies UAVs based on the obtained features. This leads to a physically interpretable machine learning algorithm that in addition to classifying UAVs is also capable of detecting UAV swarms. We demonstrate our results using data collected in indoor studio and noisy outdoor environments. Our experiments show

that our method is able to detect different types of UAVs with over 90% accuracy, even in a noisy environment. We conclude that the proposed method outperforms traditional machine learning approaches for detecting and classifying single UAVs as well as drone swarms, while retaining a high degree of interpretability. Our implementation of the proposed methods is made publicly available for reproducibility.

Keywords: Drones, wavelets, machine learning, explainability, neural networks

1 Introduction

The detection and classification of unmanned aerial vehicles (UAVs) has become an important problem in recent years. UAVs, also commonly known as drones, have been successfully applied for the delivery of goods, agriculture and various other civilian industries [1]. On the other hand, exploiting their ability to carry weaponry and surveillance equipment, UAVs have found a large array of military applications and play an ever increasing role on the modern battlefield [2–4]. In addition, drones provide a cost efficient method for criminal groups to carry out illegal activities such as smuggling [5, 6] and disrupting industry and air traffic [1, 6]. We recommend [1] and [7] for a deeper discussion on the different threats posed by malicious UAV activity.

In order to counter the threat posed by drones, the development of trustworthy and affordable technology to detect and classify UAVs is necessary. Accordingly, several recent works consider the problem of drone detection using a variety of technologies. The detection and classification of drones is a difficult measurement and signal processing task for a number of reasons. Depending on the type of the UAV, its sound, radar or visual signatures can vary widely [8]. In addition, the measurement apparatus usually needs to operate in an outside environment subject to sudden changes in weather conditions and significant background noise [8]. The appearance of so-called drone swarms, where multiple UAVs perform a coordinated task, further complicates potential defense strategies [1].

Current state-of-the-art technologies [1, 7] consider signals obtained from different types of sensors accompanied by a diverse array of signal processing methods to detect drone activity. Due to the above highlighted difficulties, it is generally accepted that a high performance UAV detection system should incorporate a variety of signals obtained from different sensors. The design of such a system can also depend on other factors, such as the type of infrastructure it is intended to protect, the overall cost of the system and specifics related to the environment (such as predictable weather conditions, noise pollution levels, etc.). Nevertheless, drone detection subsystems that depend on a single mode of measurement are an important field of study, as these provide the components of an overall drone detection system. Schematic model of the neural network used for our experiments. The samples are fed first Schematic model of the neural network used for our experiments. The samples are fed first Current UAV detection and classification technologies can be divided into four classes depending on the type of measurements they rely on. Radar based methods (see e.g., [9, 10]) use

Method	Range	Noise robustness	Universal drone detection ¹
Radar	high	high	no
Vision Based	low	low	yes
Radio Frequency	high	high	no
Sound	high	low	yes

Table 1 Different detection approaches and a comparison of their characteristics.

radio waves to detect objects. Advantages of this technology include robustness against weather conditions, long range and being able to detect the speed and direction of UAVs [1]. Unfortunately, the detection of smaller drones is a difficult task for radar systems due to small radar cross sections and low flying altitudes. In addition, radar based systems are costly due to the complexity of the involved instrumentation.

Vision based detection systems usually employ image processing to analyze recordings from cameras [11]. These are very cost effective and can provide visual confirmation of UAVs, however their performance can be hindered by visibility and weather constraints. In addition, a visual detection system is only able to detect UAVs in its line of sight.

Another interesting family of detection methods attempts to recognize drone activity by capturing and analyzing wireless signals. Methods belonging to this family include [12, 13]. Benefits of radio frequency analysis include a long operating range and the ability to classify different types of UAVs. On the other hand these methods cannot identify autonomous drones [1] or drones controlled by non-wireless means such as carbon-optic fibers [14]. Furthermore, it is important to point out the constant development of stealth technology, especially with regards to radar, and radio frequency [15].

Finally, we mention UAV detection systems that operate by recoding sound in the surrounding area [16, 17]. These systems are easy to deploy and are cost effective. Additional advantages of sound based detection schemes include the ability to classify recognized drones, estimate the position of UAVs [1] and the ability to operate without a line of sight to the target. Difficulties associated with acoustic drone detection systems include background noise filtering and vulnerability to wind conditions. A comparison of the aforementioned detection systems can be found in Table 1.

In this study, we consider a novel sound based drone detection system. We show that through the use of appropriate signal processing methods, the proposed system is able to

1. Mitigate the effect of wind and other background noise, by learning frequency signatures characteristic of UAVs of interest,
2. Classify UAV types based on the learned signatures,
3. Recognize drone swarms, that is, distinguish between intrusions by a single UAV and groups of drones.

¹Technology able to detect all drone types, even small, low-altitude vehicles, as well as state of the art UAVs, which use fiber optics.

In this study, we assume that the drones-to-be-detected, are small, electrical vehicles that are traditionally difficult to recognize [1, 7]. To reduce the cost of the system, we employ microphones based on micro-electro-mechanical systems (MEMS) technology and a mathematically justified, so-called model driven machine learning (ML) paradigm.

The use of the proposed novel signal processing scheme clearly distinguishes our approach from previous sound based UAV detection technologies. Although previous drone detectors employing artificial intelligence methods on sound signals have been introduced in [16] and [17], these rely on traditional machine learning architectures. While deep learning based approaches achieve high performance [18], model parameters do not carry physical meaning. Thus, the output of such models is not explainable. This poses concerns for safety-critical applications such as drone detection. In addition, the size of deep neural models often requires specialized hardware for real-time use, which can lead to increased deployment and maintenance costs.

In contrast, certain previous methods [16] apply static feature extraction steps to obtain meaningful information from recorded sound signals. The extracted features are then passed to different ML models, which are usually smaller than their deep learning counterparts [18]. Features are often extracted using time-frequency transformations, and thus the obtained information is interpretable. On the other hand, previous methods only considered static feature extractors. That is, these methods apply a fixed transformation (e.g., short time Fourier transform [16]) to each sound segment. This leads to sub-optimal signal representation, because the feature extractor methods cannot adapt to different UAV types, maneuvers and changing environmental conditions.

In contrast, our proposed approach relies on adaptive feature extraction transformations embedded into small neural networks. The proposed model is suitable for real time applications on limited hardware and extracts physically meaningful information from the sound signals. Furthermore it provides feature transformations that can adapt to different sound signatures and changing environmental factors.

The proposed signal processing pipeline falls into the category of so-called model driven machine learning approaches. These mathematically justified ML models are fully, or partially interpretable and can be used in safety-critical applications. Importantly, such methods retain the generalization abilities of classical deep learning approaches, at least in the context of the application to which they are deployed. A number of model driven ML methods have been recently introduced, including variable projection based neural networks [19] and kernel methods [20]. Other recent examples of model driven ML methods include wavelet convolution neural networks [21, 22].

In the current study, we propose a new wavelet convolution layer (and a corresponding model driven neural network) that uses so-called rational Gaussian wavelet (RGW) kernels [23].

RGW is a recently introduced class of admissible wavelets, whose morphology can be greatly influenced through a number of parameters. The proposed layer can learn the shape of the optimal mother wavelet, as well as a finite number of scales (corresponding to pseudo frequencies [24]) that can be used to detect and classify UAVs. It is important to point out that this adaptive quality of RGW wavelets makes them

especially useful to capture the time-frequency behaviour of the signal. In particular, the changes in the morphology of the mother wavelet make the transformation capable of separating the high and low frequency components of the signal. Therefore RGW wavelet transformation can be interpreted as an adaptive filtering technique to find relevant time-frequency components of the signal, and suppress the effect of noise caused by other vehicles, or environmental effects, such as bad weather. However, despite their adaptive morphology, RGW filters are admissible wavelets whose scale and translation parameters correspond to interpretable frequency and time components of the analysed signals.

Below, we summarize the most important novelties of our study

1. We propose a novel, RGW convolution based model driven neural network. The proposed signal processing model obtains physically meaningful wavelet coefficients from the recorded sound signals. These adaptively extracted features are then processed by a small neural network to detect UAV presence and classify UAV types.
2. We demonstrate that our proposed wavelet kernel convolution based machine learning model can be used to detect the presence of groups of UAVs.
3. We compare the proposed signal processing approach to several baseline machine learning methods. We show that in addition to the inherent interpretability of the proposed model, it significantly outperforms traditional ML methods for drone detection. All of our experiments are completely reproducible using our publicly available python implementation of the proposed methods (see data and code statement at the end of the article).
4. We propose a cost-effective, sound based system for UAV detection and classification with a light weight, interpretable ML based signal processing unit.

The rest of this paper is organized as follows. In section 2, we described the hardware used to collect our dataset and the different measurement scenarios. Section 3 provides an overview of the most important properties of the collected sound signals. These properties provide justification for the proposed signal processing pipeline. In section 4, we review some important properties of rational Gaussian wavelets [23] and introduce interpretable RGW convolution layers. These are then used to construct a small, partially interpretable neural network for drone detection. Section 5 details our experiments and results. Finally, in section 6 we summarize our findings and discuss future research directions.

2 Data acquisition

In the studio environment, an Audio/Video Recorder (Video Devices PIX 270i) received the HD-SDI signal from a wide-angle camera and simultaneously generated a 48 kHz word clock signal that was phase-locked to the video reference. Audio recordings were captured at a sampling rate of 192 kHz with a 24-bit resolution in signed Linear Pulse Code Modulation (LPCM) format. The word clock from the PIX 270i was distributed to a Master Clock Generator (Apogee Big Ben), which operated as

a frequency multiplier to derive a 192 kHz reference. The Big Ben provided this synchronized 192 kHz word clock to an Audio Interface System (Focusrite RedNet) and the associated Digital Audio Workstation (Pro Tools), ensuring a common time base across all microphone channels. The PIX 270i also recorded the video feed along with an analog reference from one of the microphones, allowing precise post-synchronization between the audio and video domains. The microphones used in the studio were Brüel & Kjær 4006 and Schoeps MK2 with CMC5 preamplifiers, all of which are pressure-type electroacoustic transducers, ensuring the most phase-coherent data acquisition possible.

To simulate real environmental conditions relevant to the intended system application, a microphone array equipped with commercial MEMS transducers was deployed outdoors. The recordings were carried out at a sampling rate of 8 kHz with a 16-bit resolution in signed LPCM format. The acoustic environment included intense background noise sources such as helicopter overflights, nearby road traffic, human speech, and strong wind. As the four microphones were mounted on a common PCB and shared identical signal-conditioning and acquisition circuitry, their outputs were recorded in temporal synchronization relative to each other.

The experiments were conducted using several commercially available multirotor UAVs, representing different size classes, acoustic signatures, and propulsion configurations. The DJI Mavic Pro and Mavic Pro 2 are foldable quadcopters with a takeoff weight of approximately 734 g and 907 g, respectively. Both platforms use four two-blade rotors with typical hover-speed ranges between 5000–7800 Rotation Per Minute (RPM).

The DJI Mavic Mini is an ultralight (249 g) quadcopter equipped with four two-blade rotors, operating at higher fundamental rotor frequencies due to its smaller propeller diameter and reduced inertia.

For outdoor measurements, additional UAV types were included, such as the DJI Mavic 3 Pro (958 g, 4 rotors), the DJI Avata 2 (cinewhoop-style ducted quadcopter, 377 g, high-RPM 3-blade rotors), and the DJI Matrice 30T, a larger industrial platform weighing approximately 3.7 kg and equipped with four large-diameter two-blade rotors rotating at significantly lower RPM ranges (2500–4200 RPM).

From Table 2 we can observe that the considered UAVs have Blade Passing Frequencies (BPF) below 200 Hz. The BPF corresponds to dividing the RPM by 60. The Nyquist-frequency implies that, in the case of an 8 kHz sampling rate, the frequency components of the signal below 4 kHz are free of distortion, meaning that the signal captures the BPF of the UAVs, as well as their initial higher-frequency harmonics. Our experiments in Section 5 demonstrate this, as our proposed detection scheme learns wavelet scales that correspond to pseudo frequencies below 600 Hz (see Fig. 7).

These platforms cover a broad acoustic spectrum: lightweight high-RPM systems produce higher-frequency harmonic structures, while heavier drones exhibit stronger low-frequency components. This diversity enables robust evaluation of the proposed signal processing and classification methodology.

UAV Model	Mass	Rotors	Blades/rotor	Typical RPM
DJI Mavic Pro	734 g	4	2	5000–7800
DJI Mavic Pro 2	907 g	4	2	5000–7800
DJI Mavic Mini	249 g	4	2	6000–8500
DJI Mavic 3 Pro	958 g	4	2	4800–7200
DJI Avata 2	377 g	4 (ducted)	3	9000–12000
DJI Matrice 30T	3.7 kg	4	2	2500–4200

Table 2 Technical specifications of the UAV platforms used in the study

3 Signal description

We review some important properties of the sound signals obtained from the front-facing Schoeps MK2 microphones described in section 2. Audio recorded from a single such microphone was used to generate the results presented in Section 5 relating to the indoor measurements, thus the input to the proposed methods are 1D signals (see Fig. 1). In the considered experiments single-channel audio signals have been used for both indoor and outdoor measurement scenarios. We note that even though the sound signals obtained from our noisy outdoor measurement scenario are worse quality, the characteristics discussed here remain true for them, which is also reflected by our experimental results (see section 5). The signals considered here have been recorded with a sampling rate of 192 kHz. The audio data is split into 100 ms long, non-overlapping segments. Thus input samples consisted of arrays containing 19200 amplitude values each. The only other preprocessing step applied to the recorded sound signals was normalization, so that the values in every segment had a mean of 0 and standard deviation 1.

As noted in [1, 16–18], the frequency profiles of the audio signals carry noticeable differences, characteristic of each drone model. This can also be seen in Fig. 2, as certain features appear consistently in the spectrogram, or the time-frequency representation obtained from the Morlet-transform of signals, only when audio from a Mavic Pro type drone is present.

The characteristics of these features, however, can be subject to change depending on factors other than the model, such as the position, or the movement of the drone. Furthermore, the measured signal segments do not exhibit periodic behavior. If the transformations used during the classification pipeline are not translation invariant, the position (in time) of the identifying features may be subject to change as well.

These qualities of the audio signals mean that drone detection is a nonlinear classification problem of nonstationary and nonperiodic signals. In section 4 we introduce an ML model based on adaptive rational Gaussian wavelet convolution operators, that can successfully process these types of signals. The experimental results of section 5 demonstrate that the scheme provides an apt solution to the identification and classification problem set in this paper.

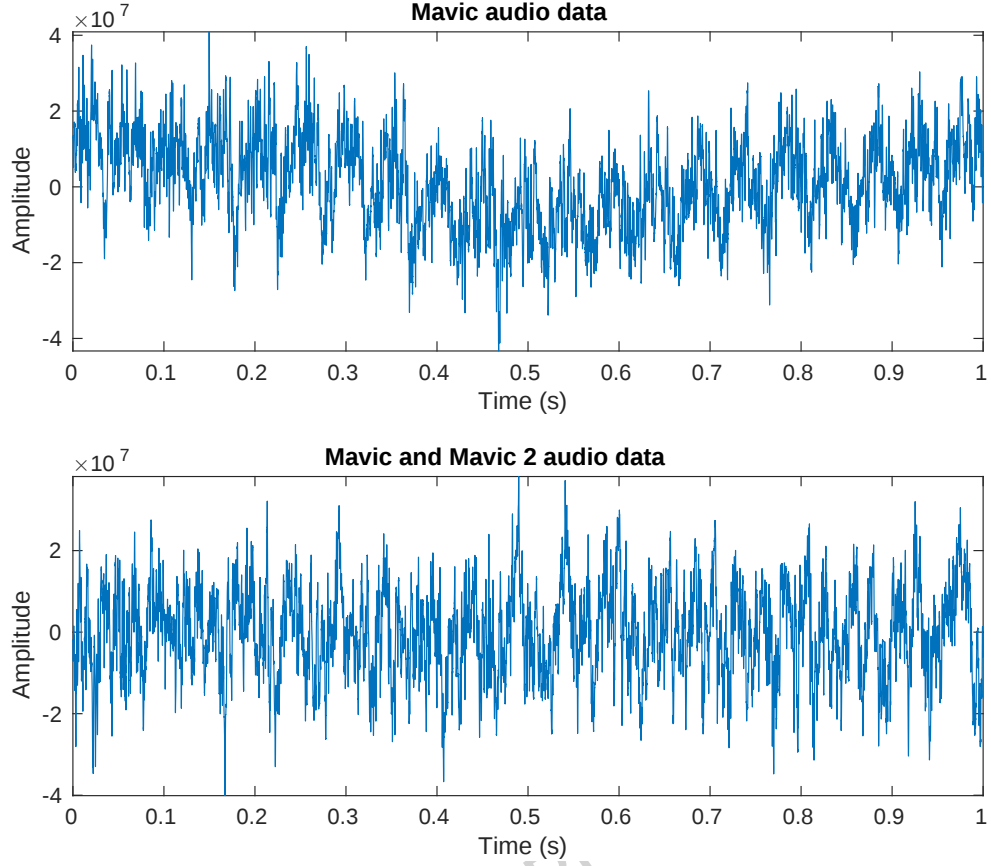


Fig. 1 Audio signals of a Mavic drone (left) and a Mavic and Mavic 2 drones flying at 1 meter high.

4 Rational Gaussian Wavelet convolution networks

Given a signal f and function ψ , the continuous wavelet transform of the signal is defined as

$$W_\psi(\lambda, \tau) = \int_{-\infty}^{+\infty} f(t) \overline{\psi_{\lambda, \tau}}(t) dt, \quad (1)$$

where

$$\psi_{\lambda, \tau}(t) := \frac{1}{\sqrt{\lambda}} \psi\left(\frac{1}{\lambda}(t - \tau)\right) \quad (t, \lambda, \tau \in \mathbb{R}, \lambda \neq 0).$$

We refer to the function ψ as the mother wavelet, while λ and τ denote so-called dilation and translation parameters, respectively. Given fixed parameters λ and τ , the number $W_\psi(\lambda, \tau)$, is called a wavelet coefficient and it describes the similarity between f and $\psi_{\lambda, \tau}$. We shall assume $\psi, f \in L_2(\mathbb{R})$ henceforth, which is a sufficient condition for (1) to exist. Furthermore, if the mother wavelet ψ satisfies the so-called

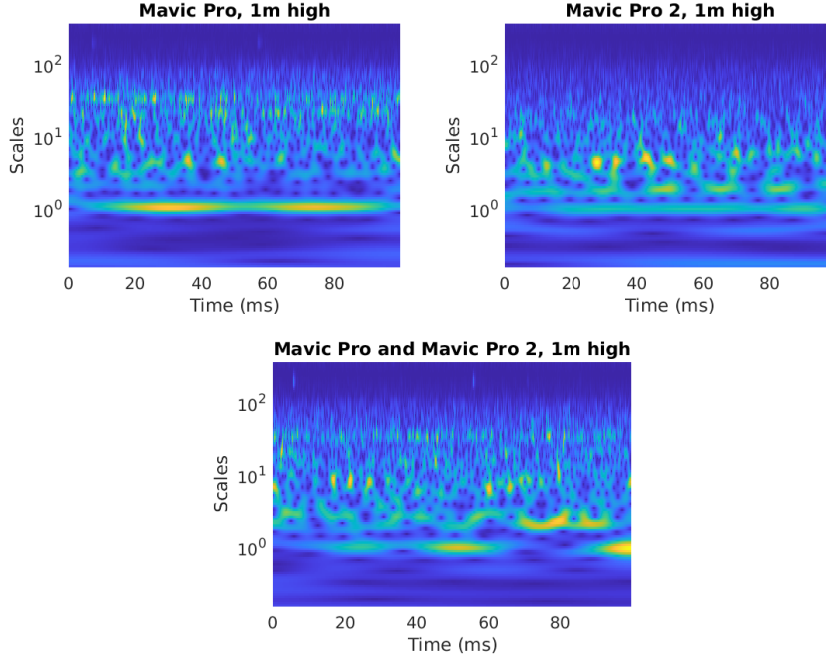


Fig. 2 Time-scale representation of audio signals of the Mavic Pro and Mavic Pro 2 drones taken indoors, with very little noise present. The top left signal is the Mavic Pro flying at 1 meter high and the top right is the Mavic Pro flying at 4 meters. The bottom is both the Mavic and Mavic 2 flying at 1 meter. The time-scale representation have been obtained via continuous wavelet transform, using Morle-wavelets.

admissibility property [24], then the transformation (1) is invertible (in the $L_2(\mathbb{R})$ -sense). The W_ψ coefficients describe the so-called time-scale representation of f which is closely related to time-frequency representations [24].

In [23] the authors introduce a family of wavelet functions, called rational Gaussian wavelets (RGW). RGWs are defined with a parameter vector $\boldsymbol{\eta} := [t_1, t_2, \dots, t_p, z_1, z_2, \dots, z_n]$ as

$$\psi^\eta(t) = C(\boldsymbol{\eta}) P^\eta(t) v^\eta(t) e^{-\frac{t^2}{2}} \quad (t \in \mathbb{R}, \boldsymbol{\eta} \in \mathbb{C}^{p+n}), \quad (2)$$

where C is a constant that depends only on $\boldsymbol{\eta}$. The rational term $P^\eta(t) v^\eta(t)$ is defined as

$$P^\eta(t) = t \prod_{k=1}^p (t - t_k)(t + t_k), \quad (t_k \in \mathbb{C} \setminus \{0\}, p \in \mathbb{N}) \quad (3)$$

and $v^\eta \in \mathcal{V}$, where

$$\mathcal{V} :=$$

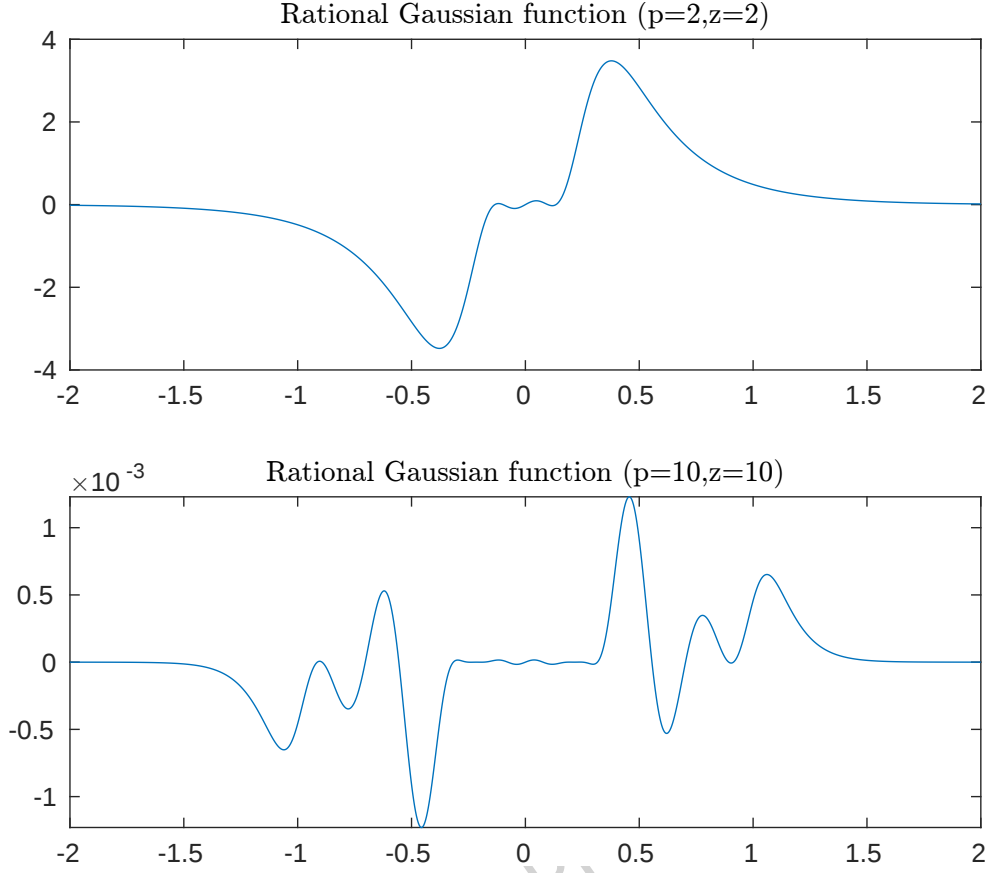


Fig. 3 Rational Gaussian functions of Eq. (2), with parameters p and z .

$$\left\{ v(t) = \frac{1}{\prod_{k=0}^{n-1} (t - z_k)(t + z_k)(t - \tilde{z}_k)(t + \tilde{z}_k)}, n \in \mathbb{N} \right\} \quad (4)$$

and

$$\tilde{z} := -\Re(z) + i\Im(z).$$

For simplicity, in this work we assume $\{t_k\}_{k=1}^n \subset \mathbb{R}$. In [23] the admissibility of ψ^η is proven, thus (in theory) f can be reconstructed from its RGW wavelet coefficients. Fig. 3 illustrates some RGW mother wavelets with different η parameter choices. Fig. 3 also demonstrates how the large number of parameters offers a high degree of flexibility for influencing the morphology of the mother wavelet. This property of RGW wavelets is also well exploited in section 5 of this study.

In [22] a wavelet based convolution kernel has been introduced, using a fixed mother wavelet, with the translation and dilation pairs as learnable parameters. The approach in [22] differs from the one introduced in this article in two regards. First, our use of RGWs allows us to optimize the morphology of the mother wavelet. Secondly, our proposed RGW convolution layer only considers the scales λ_k ($k = 1, \dots, N$, $N \in \mathbb{N}$)

as free parameters. The reason for this is that learning exact translation parameters would increase the size of the convolution layer without affecting the model's accuracy in any meaningful capacity.

An important novelty of the current study is the construction of so-called RGW convolution operators. Define the function ψ_λ as

$$\psi_\lambda(t) := \frac{1}{\sqrt{\lambda}} \psi\left(\frac{t}{\lambda}\right) \quad (t \in \mathbb{R}, \lambda > 0). \quad (5)$$

The wavelet coefficients $W_\psi(\lambda, \tau)$ from Eq. (1), where $(\tau \in \mathbb{R})$ can then be written as

$$W_\psi(\lambda, \tau) = (f * \psi_\lambda)(\tau) = \int_{-\infty}^{+\infty} f(t) \overline{\psi_\lambda}(t - \tau) dt.$$

In practical cases the signals that have to be processed are usually only available in a discretely sampled form. Let

$$f_k = f(t_k) \quad (t_k \in \mathbb{R}_+, k = 1, \dots, N),$$

and consider the notation $\mathbf{f} := (f_1, f_2, \dots, f_N) \in \mathbb{R}^N$. Let furthermore

$$\psi_{\lambda,k}^\eta := \psi_\lambda^\eta(t_k) \quad (k = 1, \dots, M)$$

denote the discrete sampling of the wavelet ψ_λ (see Eq. (5)).

Then, the wavelet coefficients of f can be approximated with the discrete convolution

$$[\mathbf{f} * \boldsymbol{\psi}_\lambda^\eta]_k := \sum_{j=1}^k \mathbf{f}_j \psi_{\lambda,k-j+1}^\eta. \quad (6)$$

Previous applications of RGW wavelets considered approximating wavelet coefficients using variable projection operators [25, 26]. This approach however heavily depends on the periodic, or quasi-periodic property of the signal. As shown in Fig. 1 and in Fig. 2, the audio signals considered in this paper are not periodic, thus variable projection based computation of RGW coefficients will not capture meaningful features. To overcome this issue, we propose discrete convolution layers with RGW kernels, that are

1. capable of extracting meaningful features from non-periodic signals using discrete convolutions,
2. allow for a high degree of adaptivity of the mother wavelet's morphology due to the nature of RGW.

The key component of the proposed model driven neural network is the RGW convolution layer. This layer applies discrete convolutions (see Eq. (6)) to the $1D$ input signals, where the kernel $\boldsymbol{\psi}_\lambda^\eta$ is an RGW wavelet characterized by a dilation parameter $\lambda > 0$ and the parameters in $\boldsymbol{\eta} \in \mathbb{R}^{p+n}$. This vector contains the zeros t_k ($k = 1, \dots, p$) of the polynomial term from Eq. (3) and the poles z_j ($j = 1, \dots, n$) from Eq. (4).

Together they greatly influence the shape of the RGW mother wavelet and can be collected to a single vector by

$$\boldsymbol{\eta} = (t_1, t_2, \dots, t_p, z_1, z_2, \dots, z_n) \in \mathbb{C}^{n+p} \\ (n, p \in \mathbb{N}).$$

The transformation characterizing the proposed RGW layer is defined by

$$\mathbf{f} \rightarrow T_{\boldsymbol{\eta}} \mathbf{f} = \begin{bmatrix} \mathbf{f} * \psi_{\lambda_1}^{\boldsymbol{\eta}} \\ \mathbf{f} * \psi_{\lambda_2}^{\boldsymbol{\eta}} \\ \vdots \\ \mathbf{f} * \psi_{\lambda_m}^{\boldsymbol{\eta}} \end{bmatrix} =: \begin{bmatrix} W_{\psi^{\boldsymbol{\eta}}}(\lambda_1) \\ W_{\psi^{\boldsymbol{\eta}}}(\lambda_2) \\ \vdots \\ W_{\psi^{\boldsymbol{\eta}}}(\lambda_m) \end{bmatrix} \in \mathbb{R}^{m \times (N+M-1)}, \quad (7)$$

where $\mathbf{f} * \psi_{\lambda_k}^{\boldsymbol{\eta}}$ ($k = 1, \dots, m$) is the discrete convolution defined in (6). The Jacobian matrix of (7), with respect to $\boldsymbol{\eta}$, can be easily computed:

$$\left(\frac{\partial T_{\boldsymbol{\eta}} \mathbf{f}}{\partial \boldsymbol{\eta}} \right)_{\lambda_k} = \mathbf{f} * \frac{\partial \psi_{\lambda_k}^{\boldsymbol{\eta}}}{\partial \boldsymbol{\eta}}. \quad (k = 1, \dots, m). \quad (8)$$

In [23], the authors offer a formula for computing partial derivatives of $\psi_{\lambda_k}^{\boldsymbol{\eta}}$ with regards to t_k and z_k . We rely on this formulation to implement the proposed RGW convolution layer.

The convolution layer is followed by a pooling layer, and a fully connected layer. The pooling layer uses a scheme in which, for every λ_k ($k = 1, \dots, m$) scale, only the wavelet coefficient of the maximal value is retained.

$$MP(T_{\boldsymbol{\eta}} \mathbf{f}) = \begin{bmatrix} \text{topQ}(|W_{\psi^{\boldsymbol{\eta}}}(\lambda_1)|) \\ \text{topQ}(|W_{\psi^{\boldsymbol{\eta}}}(\lambda_2)|) \\ \vdots \\ \text{topQ}(|W_{\psi^{\boldsymbol{\eta}}}(\lambda_m)|) \end{bmatrix} \in \mathbb{R}^{m \times Q}. \quad (9)$$

In (9), the topQ operator only keeps the $Q \in \mathbb{N}$ elements of the argument vector with the largest absolute value, and discards the rest. This allows the proposed neural network architecture to retain physically meaningful information in the following sense. For a fixed λ_k dilation, the wavelet coefficients $W_{\psi^{\boldsymbol{\eta}}}(\lambda_k)$ (see Eq. (7)) correspond to a pseudo frequency. Therefore, the output of $MP(T_{\boldsymbol{\eta}} \mathbf{f})$ can be interpreted as the amplitudes with the maximal absolute value present in the input signal $\mathbf{f}$ at the frequency band defined by λ_k . More directly, the k -th component of $MP(T_{\boldsymbol{\eta}} \mathbf{f})$ encodes the Q maximum similarity scores achieved between $\mathbf{f}$ and $\psi_{\lambda_k}^{\boldsymbol{\eta}}$ ($k = 1, \dots, m$).

The choice of Q in $MP(T_{\boldsymbol{\eta}} \mathbf{f})$ greatly impacts the structure and performance of the proposed neural network. A large Q value corresponds to a high noise filter threshold. In addition a high value of Q increases network complexity and damages interpretability. On the other hand, smaller Q values mean that the filtering of the wavelet transformation is going to be less refined. Hence, more information is lost. A demonstration of how the model's accuracy changes with Q is illustrated in Fig. 4. In our experiments Q is treated as a hyper-parameter and its value is determined using line search.

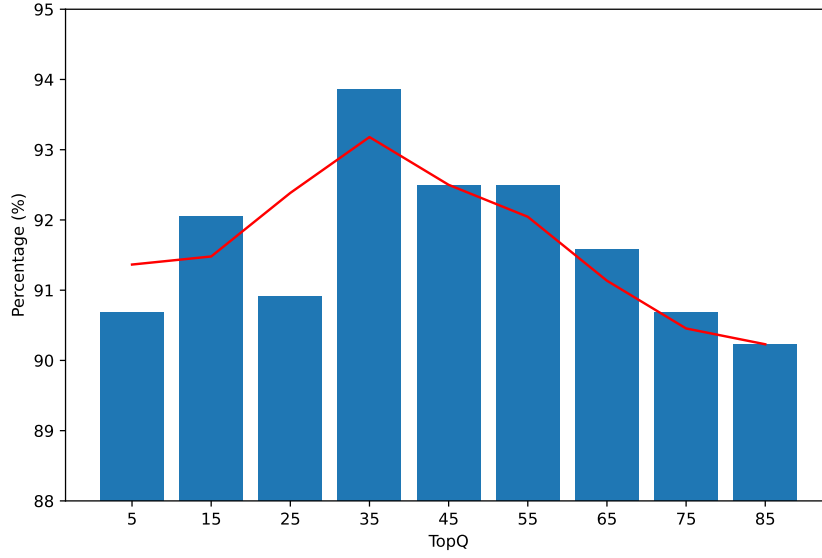


Fig. 4 The accuracy of the trained model with regards to how many values the topQ operator keeps. As the complexity of the model grows, the accuracy first improves then, after a certain number is reached, starts to decline. The displayed accuracy is always the best result of a 5-fold cross validation.

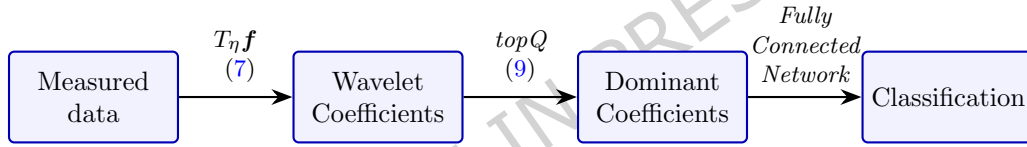


Fig. 5 A flowchart of the full pipeline. The measured data is passed through a Wavelet convolution block. The convolution output is passed to the TopQ operator which only keeps the coefficients with the maximal absolute value. Based on the dominant coefficients, the fully connected layers classify the input signal.

5 Experiments

We conducted a number of experiments to demonstrate the effectiveness of RGW convolution networks for audio-based UAV detection. As described in section 2, our dataset includes indoor and outdoor measurement scenarios along with several different types of drones and drone swarms. Accordingly, different experimental scenarios were considered for the detection and classification of a single, or multiple UAVs. The considered scenarios are described as follows.

1. Differentiating between the presence of a single UAV, and multiple UAVs. The recordings used for this experiment were taken in a studio environment. The three

Scenario id.	Task	Location	No. of UAV Types	No. of UAVs	No. of Classes
1.	Multiple UAV Detection	Indoor	3	3	3
2.	Classification	Indoor	3	1	3
3.	Detection	Outdoor	4	1	2

Table 3 Considered experimental scenarios. The number of UAVs refers to the maximum number of individual UAVs present in a single segment, while the number of UAV types refers to the total number used in that scenario.

Scenario id.	Location	No. of Categories	Segments Per Category	Segment Length
1.	Indoor	3	1790	100 ms
2.	Indoor	3	1000	100 ms
3.	Outdoor	2	1050	100 ms

Table 4 Structure of the datasets used in the individual experiments.

categories are no drones being present, the presence of a single UAV, and the presence of multiple UAVs (up to three). The UAV models used for this experiment were DJI Mavic Pro, DJI Mavic Pro 2, and DJI Mavic Mini.

2. The classification of a single UAV in a studio environment. In this case, the model receives an audio segment with a single UAV present and has to differentiate between three possible models (Mavic Pro, Mavic Pro 2 and Mavic Mini).
3. Detecting the presence of a single UAV outdoors, in a noisy environment. The categories are the same as in Scenario 1. The UAV models used for this scenario are Avata 2, Matrice 30T, Mavic Mini and Mavic Pro 3.

Table 3 gives a summary of the considered scenarios. The outdoor experiment (as described in section 2) can be regarded as extremely noisy, while the indoor experiments as noise free.

In each experiment, the recorded sound samples were subjected to an identical preprocessing regime described in section 3. The proposed ML models were very similar for each measurement scenario as well. The most significant difference, between the models used for indoors and outdoors scenarios, is the size of the used wavelet convolution kernel (7). In case the size of the kernel does not equal N , the size of the convolution output will not be $m \times N$, however it is dependant the size of the kernel and the input, and is different in each scenario. The considered neural network is composed of an RGW convolution layer, followed by two fully connected layers with ReLU activation functions before each. The output of the RGW layer is first normalized:

$$\frac{W_{\psi\eta}(\lambda_k) - \mathbf{E}(W_{\psi\eta}(\lambda_k))}{\mathbf{Var}(W_{\psi\eta}(\lambda_k))},$$

where $\mathbf{E}$ is the mean and $\mathbf{Var}$ is the variance, then downsampled with a pooling layer. The pooling layer is unique, in that it samples the dominant coefficients for each λ_k ($k = 1, \dots, m$) dilation parameter, essentially functioning as a 1-dimensional filter, as described in Eq. (9). The activation function after the final layer is either sigmoid, in the binary classification case, or softmax, in the case of multiple categories.

The number of p and n learnable parameters used to determine the morphology of the RGW mother wavelet was chosen as $p = 1$ and $n = 10$. In these experiments, we considered 10 wavelet filters (the number of scales $m = 10$). These hyperparameters have been selected by using a grid search.

The classifier block used after the RGW layer consists of a single fully connected layer with 200 neurons. The model is trained for 300 epochs with a batch size of 64.

Each individual experiment was evaluated via 5-fold cross-validation. The training consisted of randomly splitting the total training data 80-to-20 into a training set, which has been used to fit the model parameters and a test set, on which the accuracy of the model was evaluated. This process has been repeated a total of five times for the experiment, and in the following tables the mean, minimum and maximum accuracy can be seen.

Model performance has been compared to several baseline ML methods. For this purpose, we considered the Random Forest (RF) classifier [27], linear and radial Basis kernel SVM (SVM-L and SVM-RBF) classifiers [28], as well as the naive Bayes classifier [29] (NB). In addition, we conducted experiments on these scenarios using fully connected (FCNN) as well as convolutional (CNN) neural networks. In each scenario the structure of the FCNN model, and the fully connected layers in the CNN model were identical to the structure of the fully connected layers of the RGW-kernel (7) model used in that scenario. Furthermore, the number of channels, and the size of the convolution layer in the used CNN model had been identical to the one used for the RGW convolution layer. The results are shown in the following tables 5, 6 and 7.

The accuracy of the trained models are calculated by

$$Acc = \frac{N_{tp} + N_{tn}}{N_{total}}, \quad (10)$$

where N_{tp} and N_{tn} are the number of true positive and true negative prediction respectively, and N_{total} is the total number of segments in the test set.

In the following subsections we describe each experimental scenario in detail. The schematic model of the architecture can be seen in Fig 6 [R2Q3]. The flowchart of the full pipeline can be seen in Fig. 4. We note that all of the hyperparameters were determined using a large grid search of the parameter space. To ensure full reproducibility of our results, we direct the reader to the data and code availability statement at the end of the article.

5.1 Detecting swarms in a studio environment

The summary of our experimental scenarios can be found in 3. In this section we consider the indoors experimental scenario which we label with id. 1. Our dataset contains three labels: a category containing samples with only background noise, a category with sounds from a single drone, and a category which contains samples with sound recordings from multiple drones. That is, the classification task is to differentiate between a lone drone, background noise, and the presence of multiple drones.

In this particular scenario, the length of each wavelet kernel was determined to be 32, which results in the size of the convolution output being 10×19169 .

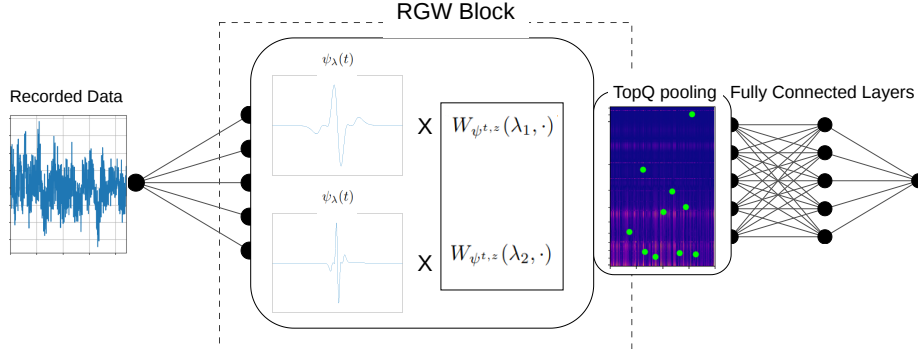


Fig. 6 Schematic model of the architecture used in our experiments. The measured data are fed first into an RGW convolution layer for feature extraction, a topQ pooling layer, and then to a fully connected block, used for classification.

The performance of the model has been compared to baseline ML algorithms, as introduced in the beginning of this section. The results can be seen in Table 5. The classical methods, as well as the fully connected neural network, fail to achieve high level of accuracy, while the RGW-kernel based network and the convolutional neural network is able to accurately identify the UAV types. Furthermore, the RGW-kernel based approach overperforms the convolutional neural network. It is important to note that the RGW-kernel based model achieves this accuracy using a lot fewer parameters and the RGW-kernel layer gives us an output whose parameters carry physical meaning about the input signals.

Model	Mean Acc.	Min Acc.	Max Acc.
WK-NN (proposed)	92.52%	90.50%	96.00%
RF	78.94%	77.47%	80.07%
SVM-L	50.32%	48.42%	51.96%
SVM-RBF	90.62%	89.20%	91.62%
NB	41.49%	39.85%	42.64%
CNN	90.57%	88.27%	92.26%
FCNN	48.13%	46.66%	49.03%

Table 5 Results of detecting the presence of a swarm in a studio environment (scenario 1. in Table 3).

5.2 Classifying drones in a studio environment

Next, we consider scenario 2. from Table 3. In the case of identifying drone model types, three categories describe our dataset. Each category contains samples from a single drone, where the categories themselves are characterized by the type of the UAV that can be heard in the sample.

The parameters of the trained model had been the same as the model used in section 5.1, with the exception of the activation function at the end being Softmax instead of Sigmoid.

The results can be seen in Table 6. Similarly to the previous scenario, the CNN and the RGW-kernel based model are able to achieve a much higher accuracy than the other approaches.

Model	Mean Acc.	Min Acc.	Max Acc.
WK-NN (proposed)	99.20%	98.33%	99.83%
RF	91.53%	90.33%	92.33%
SVM-L	38.00%	36.83%	38.83%
SVM-RBF	96.37%	95.17%	97.67%
NB	48.10%	46.33%	63.92%
CNN	99.97%	99.83%	100.0%
FCNN	47.63%	43.33%	50.83%

Table 6 Results of classifying the UAV model types present in a studio recording (scenario 2. in Table 3).

5.3 Detecting drones in a noisy environment

Finally we consider experimental scenario 3. from Table 3. For this experiment, our objective is to detect the presence of a single UAV, however the used dataset has been collected in a much less controlled environment, being subject to a large number of noise factors. The details of the noisy dataset are described in section 2.

Just like in the previous examples, the number of dilation parameters learned had been 10. The numbers p and n of RGW mother wavelet parameters remained $p = 1$ and $n = 10$, however the size of the convolution kernel was raised to 64, and the size of the convolution output changed to 10×4347 . As the task is a binary classification problem, Sigmoid had been used as an activation layer.

Table 7 contains the accuracy of the RGW-based model, as well as the accuracy of baseline ML methods, also used in the previous experiments. As can be seen, the RGW-kernel based model outperforms every other neural network, even the CNN model. It is important to note that in this scenario the task has been a lot more difficult than in the previous ones, as the audio signal was subject to heavy background noise. In this case, the interpretability of the RGW-kernel model is especially critical, as it allows us to verify the validity of the trained model.

Fig. 7 shows the output of the learned RGW convolution layer, as well as the pseudofrequency equivalent of the learned dilation parameters. On the left figure it can be observed that the learned parameters have found relevant features. The right figure show that the learned filters fall within frequency ranges associated with the BPF of the UAVs and their upper harmonics.

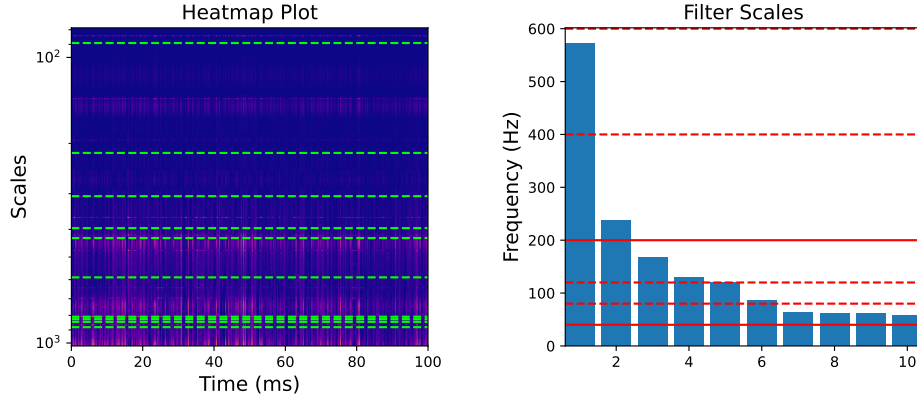


Fig. 7 The left figure shows the full RGW scalogram of an input signal, with the learned scales shown as green lines. The right figure shows the learned dilation parameters. The red lines denote the BPF of the UAVs, based on the specifications in Table 2, and the dashed lines denote the upper harmonics.

Model	Mean Acc.	Min Acc.	Max Acc.
WK-NN (proposed)	90.55%	87.75%	94.75%
RF	80.40%	78.00%	82.57%
SVM-L	62.34%	61.42%	63.42%
SVM-RBF	77.71%	74.85%	80.57%
NB	49.48%	48.85%	51.14%
CNN	89.60%	87.00%	91.25%
FCNN	55.90%	53.50%	58.75%

Table 7 Results of detecting drone presence in a noisy environment (scenario 3. in Table 3).

6 Conclusion

In this study, we introduce a novel model-based neural network, incorporating Rational Gaussian-based wavelet transformation. We have demonstrated the effectiveness of this neural network by solving a series of tasks based around acoustic detection of UAVs in various environments. By using a generic model, we have shown that our scheme can achieve accurate results, and for some tasks, can even surpass the state-of-the-art machine learning approaches, with the use of fewer parameters and without requiring large datasets.

In the future, we plan on using more specialized models in order to achieve higher accuracy on specific tasks, as well as to use these trained models for real time detection and classification. Furthermore, we plan to develop similar schemes for the purpose of not only detecting UAVs, but locating them as well. We aim to make use of our developed technologies by integrating them into more complex, multi-modal detection systems.

Acknowledgment

This research was funded by the Ministry of Innovation and Technology of Hungary from the National Research, Development and Innovation Fund, financed under the TKP2021 funding scheme, grant number TKP2021-NVA-03. TD received funding from the Swiss Government Excellence Scholarship No. 2025.0057. This work was supported by the University Excellence Fund of Eötvös Loránd University, Budapest, Hungary (ELTE). Project no. K146721 has been implemented with the support provided by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, financed under the K_23 "OTKA" funding scheme. This work was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Science. Supported by the EKÖP-KDP-24 university excellence scholarship program cooperative doctoral program of the Ministry for Culture and Innovation from the source of the National Research, Development and Innovation Fund.

References

- [1] Seidaliyeva, U., Ilipbayeva, L., Taissariyeva, K., Smailov, N., Matson, E.T.: Advances and challenges in drone detection and classification techniques: A state-of-the-art review. *Sensors* **24**(1), 125 (2023)
- [2] Bera, B., Das, A.K., Garg, S., Jalil P., M., Hossain, M.S.: Access control protocol for battlefield surveillance in drone-assisted iot environment. *IEEE Internet of Things Journal* **9**(4), 2708–2721 (2022) <https://doi.org/10.1109/JIOT.2020.3049003>
- [3] Swinney, C.J., Woods, J.C.: A review of security incidents and defence techniques relating to the malicious use of small unmanned aerial systems. *IEEE Aerospace and Electronic Systems Magazine* **37**(5), 14–28 (2022)
- [4] Wang, J., Liu, Y., Song, H.: Counter-unmanned aircraft system(s) (c-uas): State of the art, challenges, and future trends. *IEEE Aerospace and Electronic Systems Magazine* **36**(3), 4–29 (2021)
- [5] Samaras, S., Diamantidou, E., Ataloglou, D., Sakellariou, N., Vafeiadis, A., Magoulinitis, V., Lalas, A., Dimou, A., Zarpalas, D., Votis, K., *et al.*: Deep learning on multi sensor data for counter UAV applications—A systematic review. *Sensors* **19**(22), 4837 (2019)
- [6] Dedrone Inc.: Worldwide Drone Incidents. Accessed: 25 October 2023
- [7] Taha, B., Shoufan, A.: Machine learning-based drone detection and classification: State-of-the-art in research. *IEEE Access* **7**, 138669–138682 (2019) <https://doi.org/10.1109/ACCESS.2019.2942944>
- [8] Mohsan, S.A.H., Othman, N.Q.H., Li, Y., Alsharif, M.H., Khan, M.A.: Unmanned aerial vehicles (UAVs): Practical aspects, applications, open challenges, security

- issues, and future trends. *Intelligent service robotics* **16**(1), 109–137 (2023)
- [9] Wang, C., Tian, J., Cao, J., Wang, X.: Deep learning-based UAV Detection in Pulse-Doppler Radar. *IEEE Transactions on Geoscience and Remote Sensing* **60**, 1–12 (2022) <https://doi.org/10.1109/TGRS.2021.3104907>
 - [10] Tanveer, L., Misbah, M., Ahmad, A., Khan, A., Iqbal, N., Kaleem, Z.: From lab to field trials: Real-time multi-model drone detection in low snr environments. *IEEE Aerospace and Electronic Systems Magazine*, 1–10 (2025)
 - [11] Zhao, J., Zhang, J., Li, D., Wang, D.: Vision-Based Anti-UAV Detection and Tracking. *IEEE Transactions on Intelligent Transportation Systems* **23**(12), 25323–25334 (2022) <https://doi.org/10.1109/TITS.2022.3177627>
 - [12] Mandal, S., Satija, U.: Time-Frequency Multiscale Convolutional Neural Network for RF-Based Drone Detection and Identification. *IEEE Sensors Letters* **7**(7), 1–4 (2023) <https://doi.org/10.1109/LSSENS.2023.3289145>
 - [13] Nemer, I., Sheltami, T., Ahmad, I., Yasar, A.U.-H., Abdeen, M.A.R.: RF-based UAV detection and identification using hierarchical learning approach. *Sensors* **21**(6), 1947 (2021)
 - [14] Shindo, N., Kobatake, T., Masson, D., Fafard, S., Matsuura, M.: Optically powered and controlled drones using optical fibers for airborne base stations. In: *Photonics*, vol. 9, p. 882 (2022). MDPI
 - [15] Pattanaik, B., Chauhan, A.: A study of stealth technology. *Materials Today: Proceedings* **81**, 543–546 (2023) <https://doi.org/10.1016/j.matpr.2021.03.705> . International Virtual Conference on Sustainable Materials (IVCSM-2k20)
 - [16] Anwar, M.Z., Kaleem, Z., Jamalipour, A.: Machine Learning Inspired Sound-Based Amateur Drone Detection for Public Safety Applications. *IEEE Transactions on Vehicular Technology* **68**(3), 2526–2534 (2019) <https://doi.org/10.1109/TVT.2019.2893615>
 - [17] Ohlenbusch, M., Ahrens, A., Rollwage, C., Bitzer, J.: Robust Drone Detection for Acoustic Monitoring Applications. In: 2020 28th European Signal Processing Conference (EUSIPCO), pp. 6–10 (2021). <https://doi.org/10.23919/Eusipco47968.2020.9287433>
 - [18] Wang, Y., Chu, Z., Ku, I., Smith, E.C., Matson, E.T.: A Large-Scale UAV Audio Dataset and Audio-Based UAV Classification Using CNN. In: 2022 Sixth IEEE International Conference on Robotic Computing (IRC), pp. 186–189 (2022). <https://doi.org/10.1109/IRC55401.2022.00039>
 - [19] Kovács, P., Bognár, G., Huber, C., Huemer, M.: VPNet: Variable projection networks. *International Journal of Neural Systems* **32**(01), 2150054 (2022)

- [20] Dózsa, T., Deuschle, F., Cornelis, B., Kovács, P.: Variable Projection Support Vector Machines and Some Applications Using Adaptive Hermite Expansions. *International Journal of Neural Systems* **34**(01), 2450004 (2024) <https://doi.org/10.1142/S0129065724500047>
- [21] Wang, T., Lu, C., Sun, Y., Yang, M., Liu, C., Ou, C.: Automatic ECG classification using continuous wavelet transform and convolutional neural network. *Entropy* **23**(1), 119 (2021)
- [22] Li, T., Zhao, Z., Sun, C., Cheng, L., Chen, X., Yan, R., Gao, R.X.: WaveletKernelNet: An interpretable deep neural network for industrial intelligent diagnosis. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* **52**(4), 2302–2312 (2021)
- [23] Ámon, A.M., Fenech, K., Kovács, P., Dózsa, T.: Rational Gaussian Wavelets and Corresponding Model Driven Neural Networks. *IEEE Transactions on Signal Processing* **73**, 3140–3155 (2025) <https://doi.org/10.1109/TSP.2025.3592099>
- [24] Daubechies, I.: Ten lectures on wavelets. Society for industrial and applied mathematics (1992)
- [25] Golub, G.H., Pereyra, V.: The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate. *SIAM Journal on numerical analysis* **10**(2), 413–432 (1973)
- [26] Golub, G., Pereyra, V.: Separable nonlinear least squares: the variable projection method and its applications. *Inverse problems* **19**(2), 1 (2003)
- [27] Pal, M.: Random forest classifier for remote sensing classification. *International journal of remote sensing* **26**(1), 217–222 (2005)
- [28] Chapelle, O.: Training a support vector machine in the primal. *Neural Computing* **19**, 1155–1178 (2007)
- [29] Yang, F.-J.: An Implementation of Naive Bayes Classifier. In: 2018 International Conference on Computational Science and Computational Intelligence (CSCI), pp. 301–306 (2018). <https://doi.org/10.1109/CSCI46756.2018.00065>

Code and data availability

The python implementation of the proposed methods and experiments can be downloaded from

<https://gitlab.com/aele02/drone-classification-and-detection/-/tree/main>.

Data that support the findings of this study is available from HUN-REN Centre for Energy research, but restrictions apply to the availability of the data, which was used under license for the current study. The data is not publicly available.

CRedit Author statement

Gergő Ungvári: Formal Analysis, Software, Writing - Original Draft. **Ferenc Braun:** Data Curation, Writing - Original Draft. **Attila Ámon:** Software, Writing - Review & Editing. **Péter Kackstädter:** Resources, Writing - Original Draft. **János Volk:** Supervision, Funding Acquisition, Writing - Review & Editing. **Péter Kovács:** Project Administration, Writing - Review & Editing. **Tamás Dózsa:** Methodology, Software, Writing - Original Draft.

Acknowledgment

This research was funded by the Ministry of Innovation and Technology of Hungary from the National Research, Development and Innovation Fund, financed under the TKP2021 funding scheme, grant number TKP2021-NVA-03. TD received funding from the Swiss Government Excellence Scholarship No. 2025.0057. This work was supported by the University Excellence Fund of Eötvös Loránd University, Budapest, Hungary (ELTE). Project no. K146721 has been implemented with the support provided by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, financed under the K_23 "OTKA" funding scheme. This work was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Science. Supported by the EKÖP-KDP-24 university excellence scholarship program cooperative doctoral program of the Ministry for Culture and Innovation from the source of the National Research, Development and Innovation Fund.