CEAS Aeronautical Journal (2024) 15:881–893
https://doi.org/10.1007/s13272-024-00752-8

ORIGINAL PAPER

Development of an acoustic fault diagnosis system for UAV propeller blades

Leon Steinhoff[1] · Ann-Kathrin Koschlik[1] · Emy Arts[1] · Maria Soria-Gomez[1] · Florian Raddatz[1] · Veit Dominik Kunz[2]

Received: 26 February 2024 / Revised: 23 May 2024 / Accepted: 19 June 2024 / Published online: 12 July 2024

Abstract

With the rapid growth in demand for unmanned aerial vehicles (UAVs), novel maintenance technologies are essential for ensuring automatic, safe, and reliable operations. This study compares two fault detection systems that utilize the acoustic signature of UAV propeller blades for classifying their health state. By employing an acoustic camera with 112 microphones for spatial resolution of sound sources, datasets of acoustic images are generated in three differently reverberating environments for the third octave frequency bands of 6300 Hz, 8000 Hz, 10,000 Hz and 12,500 Hz. A convolutional neural network (CNN) is trained and evaluated with maximum $F1$-scores of 0.9962 and 0.9745 for two and three propeller health classes, respectively. Furthermore, we propose a second approach based on a linear classification (LC), which utilizes a rotating beamformer for comparison. This approach uses only two sound sources that are identified after the acoustic beamforming of a two-bladed propeller. In comparison, this algorithm detects propeller tip damages without applying a machine learning algorithm and reaches a slightly lower $F1$-score of 0.9441.

Keywords UAV maintenance · Machine condition monitoring · Acoustic diagnosis · Non-destructive testing · Machine learning

Abbreviations

AC	Anechoic chamber
APZ	Applikationszentrum (Application centre laboratory)
CNN	Convolutional neural network
EC	Echoic chamber
LC	Linear classifier
MRO	Maintenance, repair and overhaul
PWM	Pulse width modulation
SSH	Secure shell protocol
UAS	Unmanned aircraft system
UAV	Unmanned aerial vehicle

List of symbols

ω	Rotational speed of propeller
d	Distance between microphone array and propeller plane
f_{BF}	Beamforming frequency
PR	Peak ratio
t_i	Integration time

Units

Hz	Hertz (1/s)
m	Meters
rpm	Revolutions per minute
s	Seconds

1 Introduction

The propellers of unmanned aerial vehicles (UAVs) are prone to damage, wear and tear due to physical impacts, inappropriate handling or fatigue. Propeller blade damage can cause increased mechanical stress on UAV components, degraded performance and decreased stability. A fault detection for the propeller blades of rotary wing UAVs is, therefore, an indispensable maintenance, repair and overhaul (MRO) technology to assure airworthiness.

✉ Leon Steinhoff
leon.steinhoff@dlr.de

Ann-Kathrin Koschlik
ann-kathrin.koschlik@dlr.de

[1] Institute for Maintenance, Repair and Overhaul, German Aerospace Center, Hein-Saß-Weg 22, 21129 Hamburg, Germany

[2] Life Sciences, Hochschule für Angewandte Wissenschaften, Ulmenliet 20, 21033 Hamburg, Germany

One promising diagnostic technology is the acoustic-based condition monitoring. Installed on-ground microphones record the acoustic signature [1] which gives information about the health state of propellers [2, 3]. Previous works of Soria-Gomez et al. [4, 5] have successfully developed systems that identify the health state of the propulsion system. However, they were tested under laboratory conditions, ignoring environmental background noise, which is interfering and consequently reducing repeatability and accuracy of the diagnostics as stated by Jombo and Zhang [6].

In order to overcome this deficiency, a different data processing approach is chosen. Acoustic beamforming is an established tool that integrates a phased array of microphones to identify and localize acoustic sources, as shown by Chiariotti, Martarelli and Castellini [7]. Merino-Martínez et al. [8] show that the application of beamforming in aero-acoustics ranges from wind tunnel tests to moving aircraft source detection. The benefits of source localization is its visualization of the spatial distribution of noise sources as well as the robustness to noise [7]. The resulting visualization of the acoustics is an acoustic image which enables a new method of acoustic monitoring.

Although new electrical developments have propelled implementation of unmanned aircraft systems (UASs) in the past decade [9, 10], Martinetti, Schakel and van Dongen [11] state that maintenance is an often-overlooked topic within the large-scaled deployment of UASs. Simultaneously, the European Union Aviation Safety Agency (EASA) has issued regulations [12–14], demanding for MRO services of UASs. Consequently, reliable and cost-efficient MRO inspection technologies are desired.

Olson and Atkins [15] state the common failure modes of the motor which are caused by degradation due to prolonged use and entering particles into the housing of the motor. Therefore, fault monitoring systems are crucial for the safe operation of UAVs to inspect the integrity of the components such as actuator, sensors and main structure in order to intervene in case of an emergency. The fault diagnosis procedures are executed in real-time on-board of the vehicle. Based on these insights, the flight control laws are able to handle damages using fault-tolerant controls to adapt to the degraded system, as shown by Shraim, Awada and Youness [16], Fourlas and Karras [17], Puchalski and Giernacki [18] and Saeid, Shraim and Francis [19].

Although these methods are the current state of the art, regular maintenance is proposed in order to counteract this degradation proactively. Original equipment manufacturers (OEM) advise for visual and manual inspection prior to each flight [20]. This request becomes more challenging regarding the fact that UAV operations beyond the visual line of sight (BVLOS) with no direct and immediate access are applied more frequently. Schroth and Wackwitz [10] state, that in order to support such an operation over several flight cycles, remote landing and take-off sites are popular facilities. So-called "Vertiports" offer the opportunity to integrate automated inspection and maintenance technology, as presented by Koschlik et al. [21]. Along with the embedded fault monitoring systems on-board the vehicle and on-ground inspection, condition-based maintenance (CBM) is evaluating the actual health state of the vehicle and adapts the maintenance intervals accordingly as stated by Meissner, Meyer and Wicke [22].

Anomalous sound detection of UAV propellers with machine learning has been in the focus of research as summarized in table 1.

Jombo and Zhang [6] intensively reviewed existing methods of acoustic-based machine condition monitoring. CNNs and autoencoder (AE) are the most frequently applied machine learning algorithms, which is shown by Nunes [29] and Jombo and Zhang [6]. Soria-Gomez et al. [4, 23] suggest solutions for UAV propeller blade fault diagnosis by means of machine learning, specifically autoencoders. Liu, Chen and Zheng [27] apply a CNN, while Bondyra et al.

Table 1 Comparison of acoustic fault detection of UAV propulsion systems: background noises are not explicitly considered.

Reference	Object	Diagnosis algorithm	Accuracy	Remarks
[23]	Fixed Vehicle	Autoencoder	0.98*	Semi-anechoic chamber
[24]	Flying Vehicle	LSTM network	0.98	On-board system and outdoor testing
[4]	Single Propeller	Autoencoder	0.93*	Environmental noise
[25]	Flying Vehicle	LSTM network	0.98	On-board microphone
[26]	Single Propeller	Statistical feature extraction (DT, SVM, KNN)	0.87–0.99	Limited system & multiple algorithms
[27]	Fixed Vehicle	CNN with transfer learning	0.92	No information on environment
[28]	Fixed Vehicle	Classifier based on ANN	0.98	No background noise

Accuracies marked with * are F1-scores

[25] and Kołodziejczak et al. [24] favor a long short-term memory (LSTM) network. Altinors, Yol and Yaman [26] propose a statistical feature extraction in real-time in addition to decision tree (DT), support vector machines (SVM), and k-nearest neighbor (KNN) algorithm. Yang et al. [30] investigate the health state of the propellers based on the flight log data, applying a CNN in combination with a stack pruning sparse denoising autoencoder (sPSDAE) for noise filtering.

However, none of the above-mentioned approaches utilizes beamforming as a robust sound capturing method and all of them are ignoring parasitic noises caused by wind, ambient noises or adjacent propellers. Using acoustic imaging the environmental reverberations and noise of adjacent propellers are suppressed as parasitic noises.

This study presents the development of an acoustic fault diagnosis system for UAV propeller blades. It integrates the implementation of acoustic beamforming as a signal processing technique, paving the way towards more robust fault detection while considering the impact of background noises from different environments. We train and apply a convolutional neural network (CNN) and a linear classifier (LC) with the beamforming data under three different environmental conditions ranging from ideal anechoic to more noisy echoic environments.

Following the introduction, the state of the art of propeller diagnostics is presented. The second section describes the setup and methodology for the data generation for this use case. The third section presents the analysis of the results, which is followed by the discussion in the fourth section. In the last section, the conclusion finalizes this study.

2 Setup and methodology

An acoustic camera by CAE Systems [31] with 112 MEMS (micro-electromechanical systems) microphones and a maximal detectable frequency of 24 kHz was placed above a Holybro X500 UAV quadcopter with a maximum take-off mass (MTOM) of 1.3 kg. The UAV was mounted to a base unit (mounting plate) as depicted in Fig. 1. Two Raspberry Pi single-board computers (RP1 and RP2) were used to control the thrust of the UAV propellers. RP1 was controlled via RP2 utilizing a secure shell protocol (SSH).

While RP1 generated the pulse width modulation (PWM) signal to control the electronic speed controllers (ESC) of the UAV, RP2 acted as a WiFi hotspot. Therefore, a laptop was able to run the necessary control scripts to adjust the thrust levels of the propellers. With this setup, we managed to remotely control the UAV propellers as well as assuring reproducibility. The rotational speed of the propellers was set to approximately 4500 rpm (40 % PWM) as this is a common thrust setting for hovering.

Data from the acoustic camera was recorded on a second laptop before it was analyzed by the CNN and LC algorithms for comparison.

The distance d between the microphone array and the propeller plane was chosen to be $d = 0.28$ m. From the acoustic camera's view, the UAV was symmetrically placed in the center at position (0, 0, 0.28) as shown in Fig. 2.

The exact location of the motor was determined and a frame of 0.3 m × 0.3 m was constructed for each motor position (see Fig. 2). These frames are later used in the beamforming algorithm as the location for the acoustic images

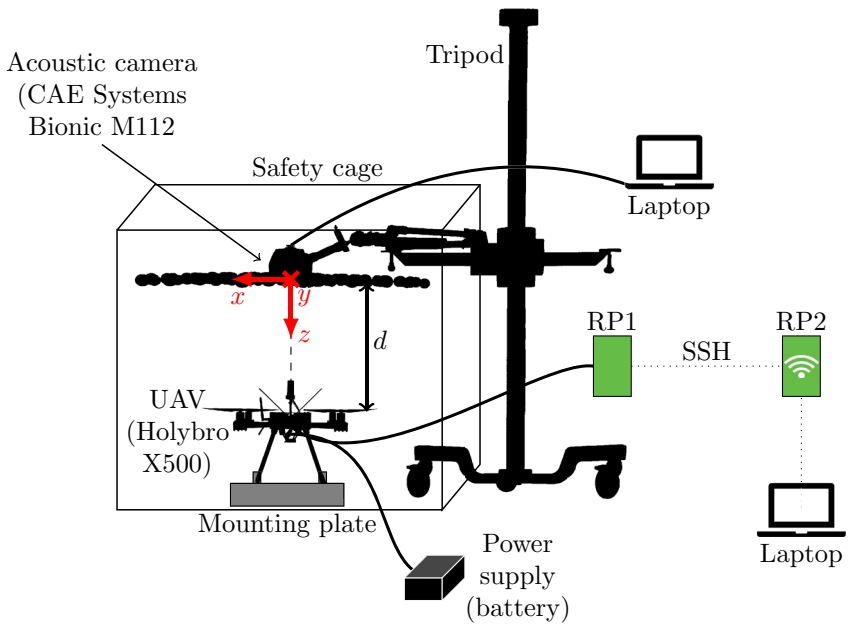


Fig. 1 Schematic overview of the experimental setup with all used devices, the used coordinate system (x, y, z) and the distance d between the microphone array and the propeller plane

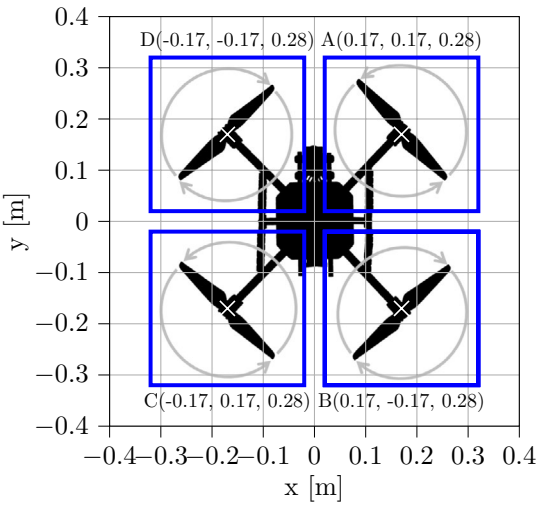


Fig. 2 UAV geometry and motor placement from the camera's view, where A(0.17, 0.17, 0.28) are the Cartesian coordinates (x, y, z) of the motor position of motor A, B, C, and D accordingly

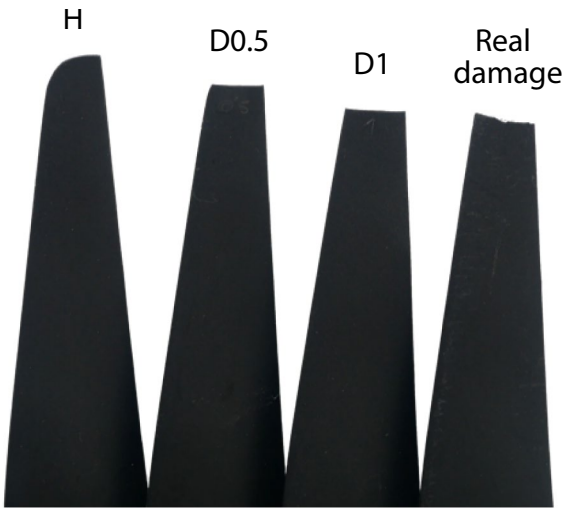


Fig. 3 Blade classes with healthy (H) and damaged classes, where the blade tip is cut off by 0.5 cm (D0.5) and 1 cm (D1). Furthermore, a real damage is included. All propellers are double-bladed with damages located on only one blade

and, therefore, isolating a single propeller from the rest of the UAV.

2.1 Generation of training data

The CNN is trained and evaluated using the Python open-source framework TensorFlow [32] and Keras. Data generation for the CNN involved the quadcopter UAV recorded with the acoustic camera in three different environments. They are abbreviated as follows:

- Application center laboratory (**APZ**): a regular laboratory environment with concrete walls and a large window front.
- Anechoic chamber (**AC**): an acoustic chamber with fully anechoic characteristics. Here, the mounting plate was removed because of its reflective surface.
- Echoic chamber (**EC**): an acoustic chamber with fully echoic characteristics.

Furthermore, three different propeller configurations depicted in Fig. 3 including undamaged and damaged propellers were tested. In this paper, the following abbreviations are used to indicate the health classes of a propeller blade:

- "**H**" stands for healthy (i.e., undamaged) blades.
- "**D0.5**" stands for damaged blades with a cut tip of 0.5 cm length (approximately 2 % of the blade length).
- "**D1**" stands for damaged blades with a cut tip of 1 cm length (approximately 4 % of the blade length).

For every recording of the UAV, all motors (A, B, C and D) were running continuously. To avoid repeating patterns in the distribution of damaged and undamaged propellers among the four motors, all distributions, which are not symmetrical to each other, have been recorded with specific recording times. Table 2 lists the distributions.

We obtain a recorded time for every class and environment by summing the recording times for all individual classes in the three environments. They are shown in Table 3.

To avoid a class imbalance, data for class "H" in the anechoic chamber was limited to 1200 s. We split the data into three different sets: the training dataset, the validation dataset and the testing dataset. The training dataset is the biggest portion of the original data with 70%. It is used to train the models, i.e., adjust the weights of the CNN. 10% of the original data is used for validation during the training. This step is important for evaluating the model's accuracy with data it was not presented with so far. The remaining 20% are used for testing. The testing data can be interpreted as a dataset that is used after the training has stopped to evaluate the final state of the network. The distribution of propeller blade classes among all datasets was done randomly.

2.2 Data processing

Figure 4 illustrates the data processing flow of the CNN approach (left) and the LC approach (right).

The input data was converted from the output data format of the acoustic camera (.tdms) to the H5 hierarchical data format (.h5) and split into files with a duration of 1 *s* before the beamforming algorithm was applied. For all

Table 2 Overview of propeller distributions among motors A–D with their respective recording time in all three environments

Distribution				Recorded time [s]		
A	B	C	D	APZ	AC	EC
H	H	H	H	60	120	60
D1	H	H	H	120	120	120
D1	D1	H	H	120	120	120
D1	D1	D1	H	120	120	120
D1	D1	D1	D1	60	60	60
H	D1	H	D1	120	120	120
D0.5	H	H	H	–	120	–
D0.5	D0.5	H	H	–	120	–
D0.5	D0.5	D0.5	H	–	120	–
D0.5	D0.5	D0.5	D0.5	–	60	–
H	D0.5	H	D0.5	–	120	–

Table 3 Total recorded time for the three classes H, D1 and D0.5

Class	Recorded time [s]		
	APZ	AC	EC
H	1200	2400	1200
D1	1200	1200	1200
D0.5	–	1200	–

beamforming algorithms, four different frequencies were evaluated: 6300 Hz, 8000 Hz, 1000 Hz and 12500 Hz, each with a third octave frequency band.

As shown by Dougherty [33], the functional beamformer is well suited for aeroacoustic sound sources. It was implemented with the open-source Python library Acoular [34]. The resolution of the acoustic images is $50 \times 50 \times 1$. After the images were created, they were normalized to meet the required range of 0 to 1 for the CNN input. As a result the dynamic range is maximized, but quantitative

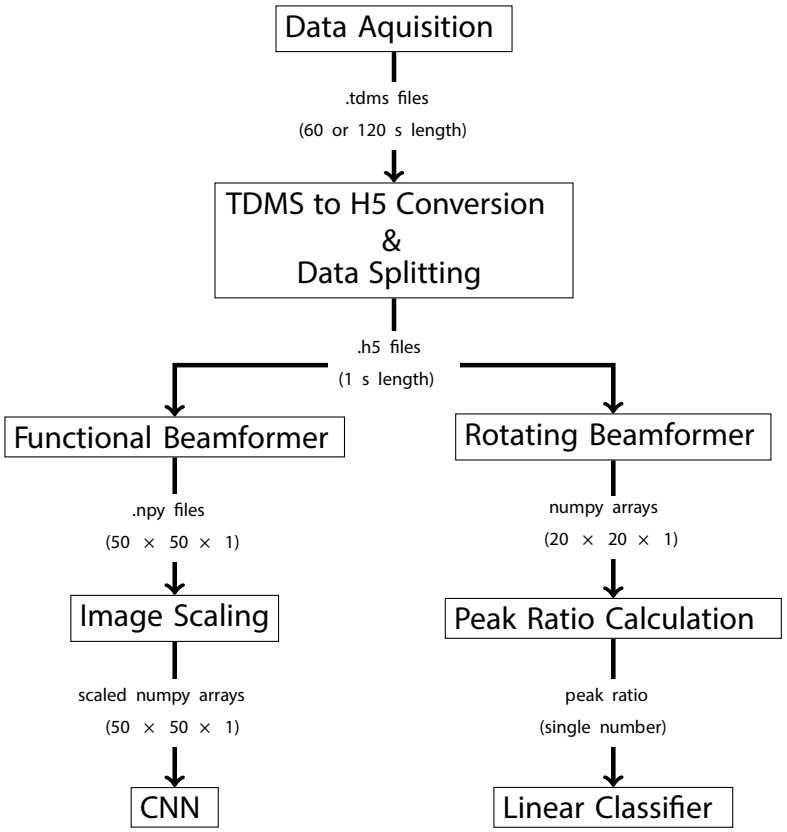


Fig. 4 Flow diagram of the data processing from the recording to the classification system input

information about the volume of the acoustic source is lost. The normalized images are then processed by the CNN. This specific neural network is chosen because its convolutional nature is well suited for processing images as inputs.

The rotating beamforming algorithm rotates the acoustic map at the same speed as the rotating source around the center of rotation of the source, as shown by Sijtsma [35]. This technique eliminates the rotary movement of the source and enables a true acoustic map that resembles the source as if it was stationary. For the accurate calculation of a rotating beamforming image, the rotational speed ω of the source must be known precisely. There was no tachometer used during the data acquisition. We, therefore, developed an algorithm based on the fluctuation of brightness patterns to predict the true ω of the propeller. The LC was applied as an alternative that uses no machine learning. It classifies the rotating beamforming images on the basis of just one feature. This benefits the computation time and the overall system's complexity.

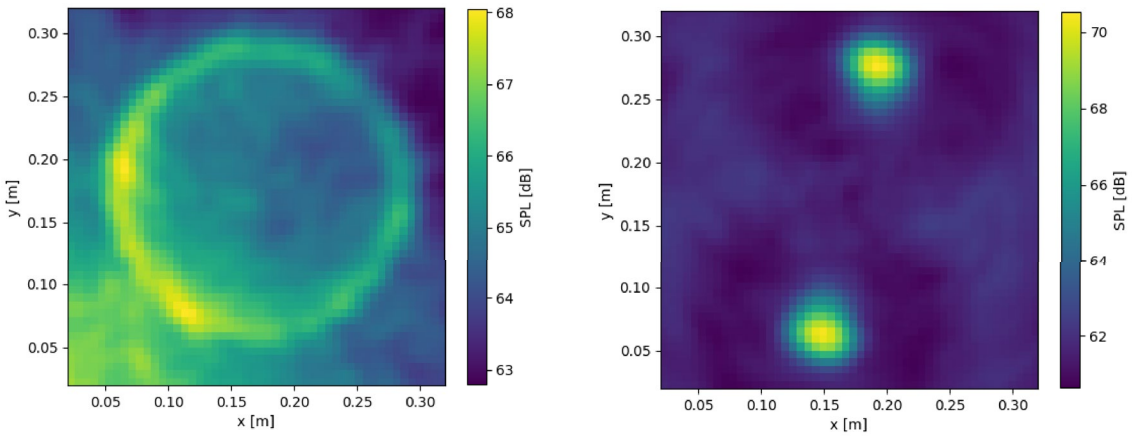
Figure 5 compares functional and rotating beamforming of a two-bladed propeller as a sound source. f_{BF} is the center frequency of the frequency band. t_i is the integration time of the beamforming. An extensive review on beamforming algorithms, their use-cases and results is given by Merino-Martínez et al. [8].

2.3 Hyper-parameter gridsearch

There are several hyper-parameters that characterize CNNs. Determining the accuracies of neural networks trained with different combinations of hyper-parameters is essential for maximizing the classification performance. There are different methods for determining the optimal hyper-parameters. We chose a gridsearch algorithm, where all combinations of predefined hyper-parameter values are compared. This algorithm is beneficial for determining the global trend of the influence of a hyper-parameter. Local optimizers, e.g., the Adam optimizer, could later be used for finding local minimums and fine-tuning. The following hyper-parameters were chosen for the gridsearch algorithm:

- **Architecture:** 3 numbers are defining the dimension of the different layers of the network. The first two numbers are the depth of the first two convolutional layers, respectively. The third number is the dimension of the first fully connected layer. The chosen values for the architecture in the gridsearch were: 8-8-32, 16-16-16, 8-16-16, 16-8-32, 4-4-64, 8-4-64, 4-8-32, 4-16-16 and 16-4-64. The values were chosen in such a way, that the resulting trainable weights of the CNN were not larger than the total number of available input data. The total amount of input data summed up to 14,400 files per class.
- **Dropout:** Three dropout layers were included in the CNN architecture. Here, dropout refers to the percentage of randomly distributed zero weights in each dropout layer to prevent over-fitting. We chose 0 %, 10 %, 20 % and 30 %.
- **Learning rate:** the learning rate corresponds to the step size of the learning algorithm. The selected learning rate values for the gridsearch were 0.0005, 0.002 and 0.005.

Nine different architectures, 4 different dropout values and 3 different learning rates result in a total of 108 possible combinations of the hyper-parameters. To account for



(a) Functional Beamforming (for CNN), $f_{BF} = 10000\,\mathrm{Hz}$, third octave, $t_i = 1\,\mathrm{s}$

(b) Rotating Beamforming (for LC), $f_{BF} = 10000\,\mathrm{Hz}$, third octave, $t_i = 1\,\mathrm{s}$, $\omega = 75.49\,\frac{1}{s}$

Fig. 5 Comparison between Functional Beamforming and Rotating Beamforming. Note, that the images have just one value per pixel and thus should be shown in gray-scale. For a better visualization of the dynamics of the images, a color-scale was chosen

statistical variance in the training process, each parameter combination is repeated 3 times. The total number of CNNs that were trained is, therefore, 324. To compare the classification performance of the parameter combinations, the average validation accuracy of the 3 repetitions is calculated. For all training processes, early stopping was used. This method is used for controlling the amount of training epochs of a neural network. The validation loss was the training metric that was monitored. If the validation loss would not decrease after 5 consecutive epochs, the training process would be stopped. The maximum number of epochs was set to 60.

3 Results

This section presents the experimental results and the comparison between the two proposed approaches, the CNN and the LC approach. For comparing the accuracy of classification models, a harmonic mean of the precision and recall is commonly applied. This harmonic mean is called the F1-score, proposed by Chinchor [36].

3.1 Results of the CNN model

Figure 6 illustrates the comparison between images with different frequency bands as inputs for the CNN models and their resulting averaged accuracy. The F1-score reaches higher values with higher frequency bands. The reason for the better performance is the increasing sharpness of the acoustic images with increasing frequency due to a smaller beamwidth of the corresponding point spread function (PSF). The higher the frequency, the sharper and visually more defined are the acoustic sources in the image.

Due to the higher classification accuracies, all the models in the following sections were based on 12500 Hz frequency band input data. The speed of evaluation did not vary among the different input data.

The results of the gridsearch (see section 2.3) with data from the different environments is presented in the following subsections.

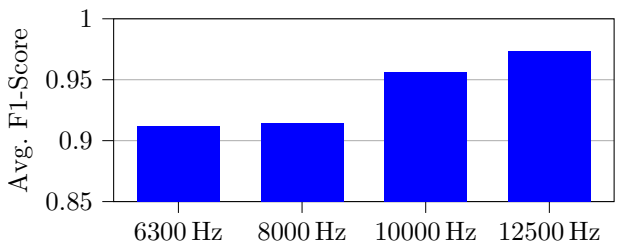


Fig. 6 Comparison of the average F1-scores of CNN models trained with different frequency bands as input data

3.1.1 Anechoic chamber dataset-2 classes

The results of the gridsearch based on the data from the AC are illustrated in Fig. 7.

Low dropout rates are beneficial with an average validation accuracy of almost 0.99 at 0% dropout. The learning rate is well scaled around 0.002. The architecture shows a slightly better performance for models with a larger fully connected layer, which is specified in the last number of the architecture.

Of all 324 CNN models, the one with the highest validation accuracy is specified in Fig. 8. With a layer architecture of 4-16-16, no dropout and a learning rate of 0.005, it reaches a validation accuracy of 0.9972 and an $F1$-score of 0.9943 on the test data. The corresponding confusion matrix of the test data is also depicted in Fig. 8.

The learning curves of the accuracy and loss measures are shown in Fig. 9. The early stopping is reached after 23 epochs.

3.1.2 Laboratory Dataset-2 Classes

The gridsearch for the APZ data, shown in Fig. 10, shows the same behavior of the dropout as with the AC environment data. The learning rates of 0.0005 and 0.002 produce a slightly higher accuracy than the higher value of 0.005.

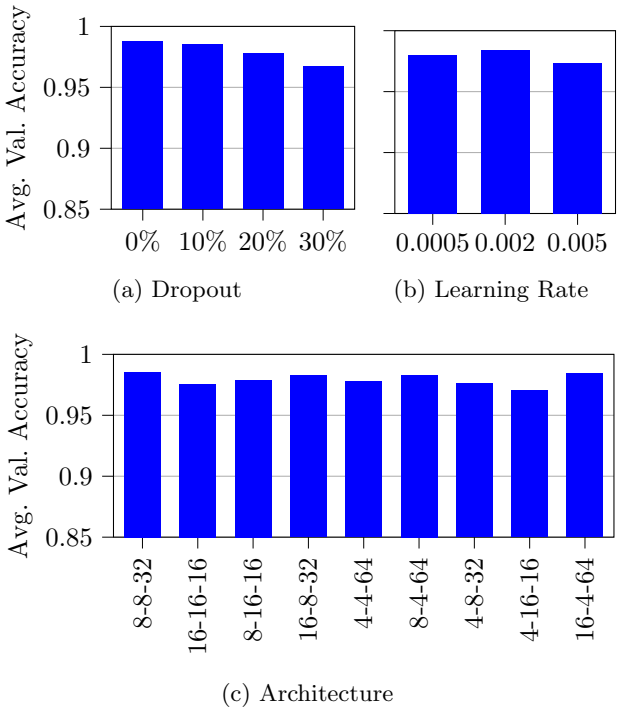
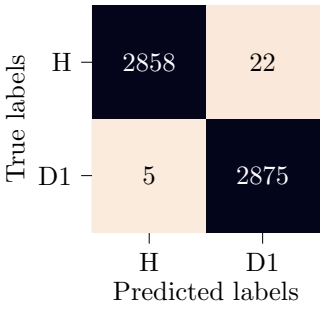


Fig. 7 Gridsearch results of the AC 2-class classifier showing the average validation accuracy over the hyper-parameters dropout (a), learning rate (b) and architecture (c)

Fig. 8 Parameters and confusion matrix of the best CNN model from the AC 2-class classifier gridsearch

Architecture	4-16-16
Dropout	0 %
Learning rate	0.005
# of weights	31,641
Epochs	24
Val. accuracy	**0.9972**
F1-score	**0.9943**



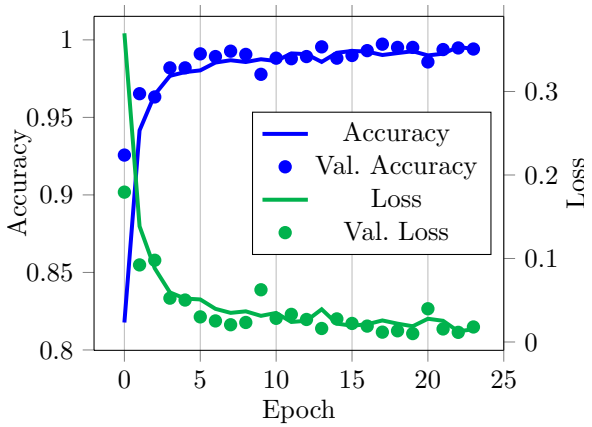


Fig. 9 Training and validation accuracy and loss of the best CNN model from the AC 2-class classifier gridsearch

There is no significant correlation between the architecture and the classification performance in this environment.

The model with the highest validation accuracy is presented in Fig. 11. It reaches a higher accuracy compared with the model in the AC. It used 42 training epochs, no dropout and a low learning rate. Figure 11 also shows the confusion matrix with very few falsely classified images. The learning curves are shown in Fig. 12

3.1.3 Echoic chamber dataset-2 classes

For the EC, a low dropout value is beneficial. The learning rate influence is similar to the gridsearch in the APZ, where 0.0005 and 0.002 are the best performing learning rates. Again, there are no dominant correlations between the layer architecture and the average validation accuracy. Figure 13 shows the respective bar plots.

Figure 14 presents the best performing CNN model with a validation accuracy of 0.983. It has similar parameters to the previous gridsearches. The training process stretched over 38 epochs.

Figure 15 shows the learning curves of best CNN model.

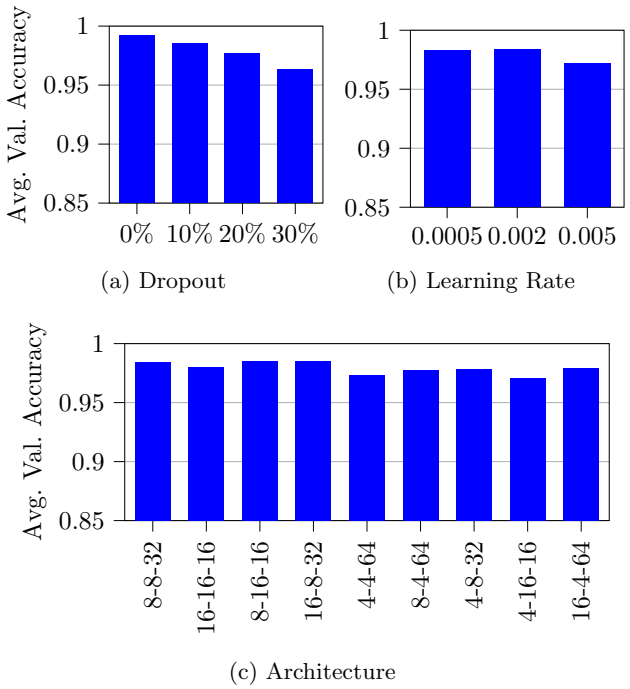


Fig. 10 Gridsearch results of the APZ 2-class classifier showing the average validation accuracy over the hyper-parameters dropout (**a**), learning rate (**b**) and architecture (**c**)

3.1.4 Anechoic chamber dataset-3 classes

The introduction of the third propeller state "D0.5" (see Fig. 3) to the training data does not change the influence of the different hyper-parameters compared to the 2-class CNN models. The dropout rate is again showing a clear tendency towards lower values as well as the learning rate. The CNN models with larger fully connected layers are performing slightly better, similar to the effect with the 2-class AC data (see Sect. 3.1.1). Figure 16 shows the results.

A CNN model with an architecture of 8-16-16, 0% dropout and a learning rate of 0.002 is the optimal configuration for the highest accuracy among all 324 neural networks. The models of the 3-class classification generally have a lower average accuracy and F1-score than their 2-class counterparts. The biggest weakness of the model is to correctly

Architecture	8-16-16
Dropout	0 %
Learning rate	0.0005
# of weights	32,257
Epochs	42
Val. Accuracy	**0.9986**
F1-score	**0.9977**

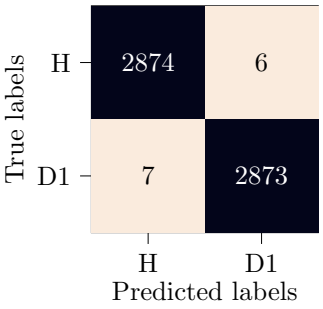


Fig. 11 Parameters, results and confusion matrix of the best CNN model from the APZ 2-class classifier gridsearch

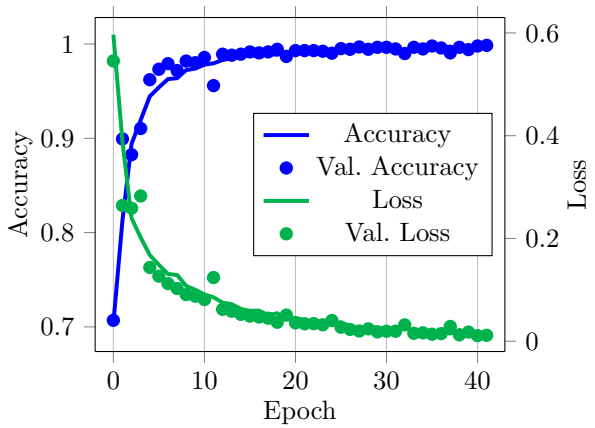


Fig. 12 Training and validation accuracy and loss of the best CNN model from the APZ 2-class classifier gridsearch

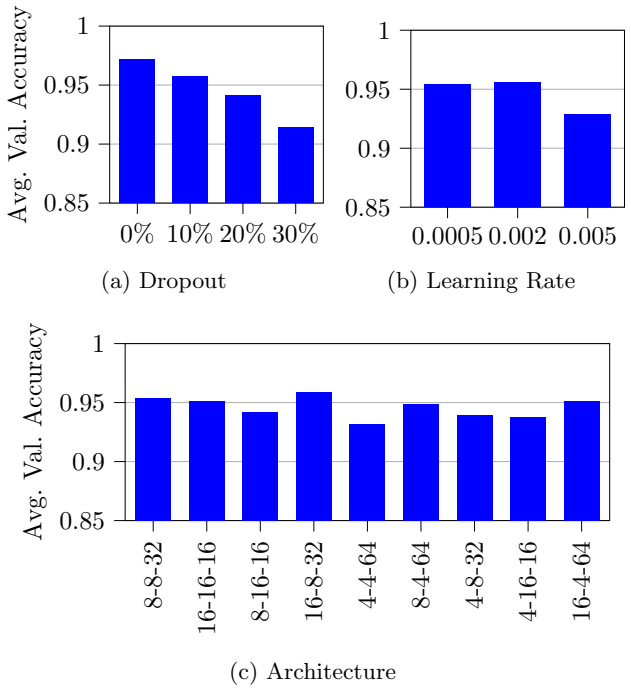


Fig. 13 Gridsearch results of the EC 2-class classifier showing the average validation accuracy over the hyper-parameters dropout (**a**), learning rate (**b**) and architecture (**c**)

identify the "D1" propellers, which can be seen in the confusion matrix in Fig. 17.

Figure 18 provides an overview of the learning curves for the above-mentioned best CNN model indicating a normal progression.

3.2 Results of the LC model

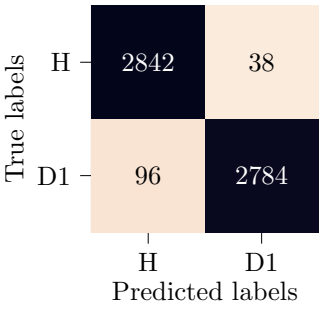
The linear classifier relies on the ratio between the maximum of the two sources in a rotating beamforming image, where a sound source is shown for every blade tip (see Fig. 5). As the sources are shown as peaks in the image, the ratio is called the "peak ratio" *PR* from here. It was found, that propellers with one healthy and one damaged blade tip have a higher *PR*, than propellers with two healthy blade tips. This approach does not need a machine learning model but instead, diagnoses the propeller's classes (H, D0.5 or D1) with extensive post-processing of the recorded data. By plotting a histogram of the peak ratios with regard to the corresponding class, it is possible to determine a threshold above which the image is classified as damaged. If the image has a *PR* below the threshold, it is classified as undamaged. Equivalently, images with a *PR* above the threshold are classified as damaged.

The histogram in Fig. 19 depicts the distribution of the peak ratios *PR* with respect to the classes.

With the introduction of a threshold, a LC model can be modeled and tested with rotating beamforming images. Similarly to the CNN approach, the available data were split into training data to determine the best threshold and testing data for evaluating the performance of the threshold with data that were not involved in determining the threshold. A validation dataset was not needed, as there was no validation during the training. The training data were 70% and the testing data 30% of the original dataset size. Due to the randomly shuffled dataset before the splitting into training and testing data, the accuracy of classification varied between individual training runs. We, therefore, performed 10 iterations of the threshold determination for each environment and the average $F1$-score and threshold were treated as the result.

Fig. 14 Parameters, results and confusion matrix of the best CNN model from the EC 2-class classifier gridsearch

Architecture	16-16-16
Dropout	0 %
Learning rate	0.002
# of weights	33,489
Epochs	38
Val. Accuracy	**0.9830**
F1-score	**0.9767**



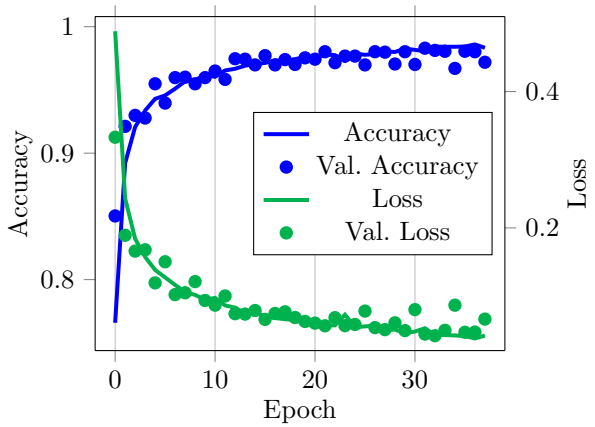


Fig. 15 Training and validation accuracy and loss of the best CNN model from the EC 2-class classifier gridsearch

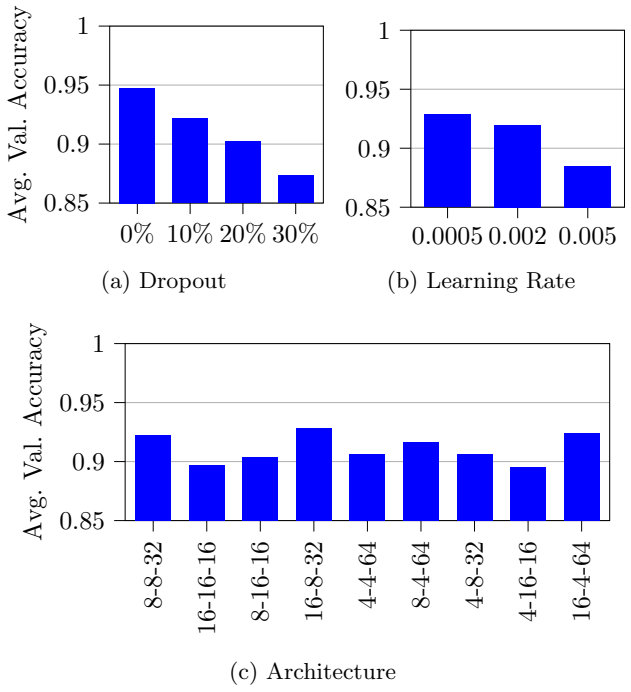


Fig. 16 Gridsearch results of the AC 3-class classifier showing the average validation accuracy over the hyper-parameters dropout (**a**), learning rate (**b**) and architecture (**c**)

The average $F1$-score was 0.9441. The highest $F1$-score of a single model was 0.9686, which can compete with some of the CNN models that were tuned with the gridsearch. Figure 20 shows the threshold, number of input data and the corresponding confusion matrix. The threshold is also depicted in Fig. 19.

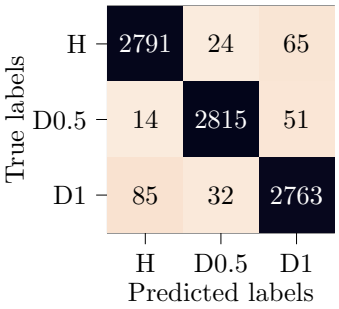
When adding the "D0.5" health class, the LC needs two thresholds to separate the three clusters. The performance drops to an F1-score of 0.6366 because the data of three classes has more overlaps in the *PR* distribution and is, therefore, less separable.

4 Discussion

The objective of overcoming the parasitic reverberation effects in acoustic condition monitoring was met with both approaches due to the inclusion of beamforming algorithms. In order to demonstrate the potential, three datasets were created, representing both ideal and realistic scenarios. The validation accuracy of the CNN reaches an F1-score of 0.9972 for the AC dataset, which drops to 0.9830 for the highly reverberating dataset, which is a performance difference of 1.4 %. The previous studies of Soria-Gomez et al. [23] without beamforming and multiple propeller inspection were able to achieve an average $F1$-score of only 0.88. Although the authors of [23] were able to improve their accuracy results for a single propeller inspection to 0.98, the process time is longer due to the successive inspection routine of the propellers. Based on this comparison, the presented algorithm shows improved results in terms of quality and process time. Another advantage of the proposed method over the fault diagnosis of [23] is the ability of localizing the damage and quantifying the extend with the multi-class neural network. The maximum $F1$-score of the proposed system by Bondyra et al. [25] is 0.985, which is slightly lower than the best CNN models from the AC and APZ environments. The propeller fault detection described in Ref. [25] was implemented as an on-board solution, while the proposed method in this research does not add any additional weight to the UAV. Altinors, Yol and Yaman

Fig. 17 Parameters, results and confusion matrix of the best CNN model from the AC 3-class classifier gridsearch

Architecture	8-16-16
Dropout	0 %
Learning rate	0.002
# of weights	32,257
Epochs	31
Val. Accuracy	**0.9725**
F1-score	**0.9686**



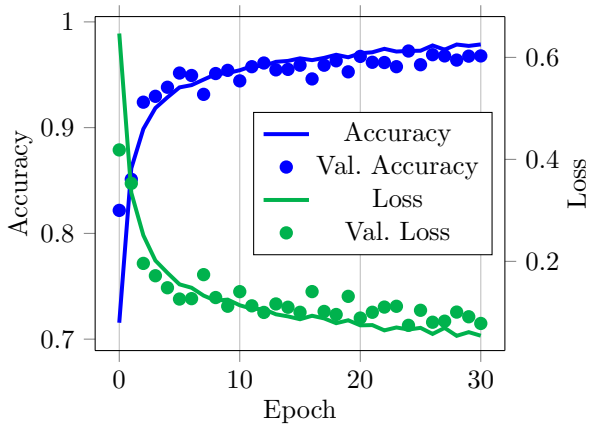


Fig. 18 Training and validation accuracy and loss of the best CNN model from the AC 3-class classifier gridsearch

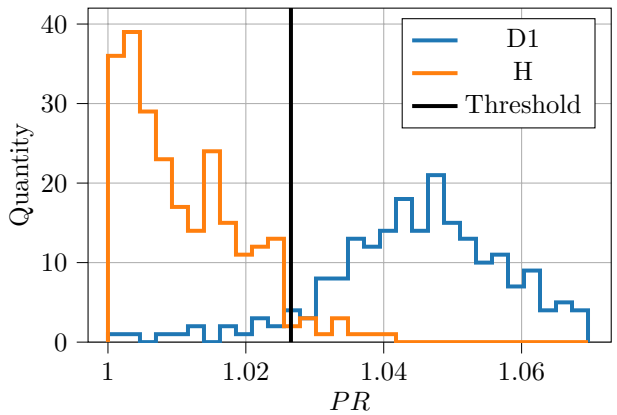


Fig. 19 *PR* distribution and best threshold for the LC demonstrated for EC environment with two classes. A clear division of the data clusters with only a small overlap can be observed

[26] reach similar F1-scores of up to 0.998 but only for a single propeller blade and without quantification.

The damage type we focused on in this work was a cut blade tip. Randomly damaged propellers that would appear on UAVs operated in a real environment, are not represented in the training data. Therefore, it cannot be excluded that over-fitting to the specific damage type is present in the CNN and LC models. The research can be extended with additional recordings of a more diverse dataset with different UAV positions, distances d in the experimental setup and different damage types. This would prepare the system for a real-world scenario, where the UAV would always be in slightly different positions and would have different propeller damage types. The diversification of the training data would increase the robustness of the classification. This can also be achieved using transfer learning.

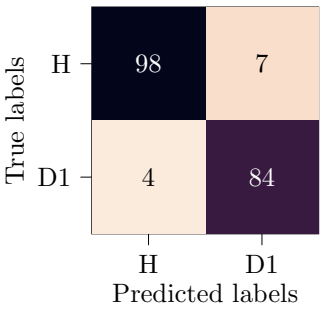
Similar results were achieved by the application of the LC algorithm based on the peak ratios of the rotating beamforming. With an accuracy of 0.9441, this approach is slightly less accurate as the CNN models. However, the true strength of this algorithm lies in its physical interpretation. Soria-Gomez et al. [23] use a similar approach and propose a threshold based on the rotational speed and the blade passing frequency (BPF). However, this algorithm is limited to a bandwidth of 4 kHz and fails when multiple propellers are inspected simultaneously. The fault diagnosis system in this study is more robust. We demonstrated that the focus on higher frequency bandwidths, e.g., 12 kHz, is beneficial for acoustic monitoring of UAV propeller blades. Furthermore, the linear classifier does not require a neural network and is computationally less expensive.

5 Conclusion

The proposed acoustic fault diagnosis system for UAV propeller blades shows very promising results with both, the CNN approach and the LC approach. The measured sound emission at the tips of rotating propeller blades contains sufficient information about the health class of the propeller tip and this information can be reliably extracted with the proposed methods. For the machine learning approach, extensive and diverse datasets are necessary to train a CNN model for the real-world application. Our study shows that acoustic images calculated with the functional beamformer are a suitable input for a CNN especially for higher frequency bands, because the sound source identification is more accurate when using small beamwidths. When comparing the results with the literature for on-ground

Fig. 20 Results and confusion matrix of the EC 2-class LC

# of input data	645
Avg. Threshold	1.0265
Avg. F1-score	**0.9441**



fault diagnosis system, the F1-score for simultaneous inspection of all propeller blades is improved by 10 %. The LC approach does not require machine learning making it an interesting alternative. However, the performance drops substantially when the LC is required to identify more than two health classes.

Author Contributions Conceptualization, L.S. A-K.K.; methodology, L.S., A-K.K.; software, L.S., E.A.; validation, L.S., E.A., A-K.K.; formal analysis, L.S., E.A., A-K.K.; investigation, L.S., A-K.K.; resources, F.R.; data curation, L.S., M.S-G.; original draft preparation, L.S.; writing, review and editing, L.S., E.A., A-K.K., V.D.K.; visualization, L.S., E.A.; supervision, A-K.K., V.D.K.; project administration F.R., A-K.K. All the authors have read and agreed to the published version of the manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data Availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors did not receive support or funding from any organization for the submitted work.

References

1. Száka´l, A.: ed. CINTI 2015: 16th IEEE International Symposium on Com394putational Intelligence and Informatics : proceedings : 2015, November 19-21,395 Budapest. Piscataway, NJ: IEEE, (2015). isbn: 978-1-4673-8520-6
2. Pechan, T., Sescu, A.: Experimental study of noise emitted by propeller's surface imperfections. Appl. Acoust. **92**, 12–17 (2015). https://doi.org/10.1016/j.apacoust.2014.11.014
3. Semke, W.H., Zahui, D.-K., Schwalb, J.: The vibration and acoustic effects of prop design and unbalance on small unmanned aircraft. In: 2191-5652, pp. 9–16. (2021). https://doi.org/10.1007/978-3-030-47713-4_2
4. Soria-Gomez, M., Koschlik, A.-K., Arts, E. , Raddatz, F.: Non-destructive evaluation of the condition of a UAV's propellers by means of acoustics. In: NDE 4.0, Predictive Maintenance, and Communication and Energy Systems in a Globally Networked World. Ed. by Norbert G. Meyendorf, Christopher Niezrecki, and Saman Farhangdoust. SPIE, 6.03.2022 - 11.04.2022, p. 19. ISBN: 9781510649736. https://doi.org/10.1117/12.2612770
5. Soria-Gomez, M., Koschlik, A.-K., Arts, E., Raddatz, F., Wende, G.: Acoustic non-destructive testing of UAS's propellers during predeparture and post-flight checks. Res. Rev. J. Nondestruct. Test. (2023). https://doi.org/10.58286/28093
6. Jombo, G., Zhang, Y.: Acoustic-based machine condition monitoring-methods and challenges. Eng **4**(1), 47–79 (2023). https://doi.org/10.3390/eng4010004
7. Chiariotti, P., Martarelli, M., Castellini, P.: Acoustic beamforming for noise source localization—reviews, methodology and applications. Mech. Syst. Signal Process. **120**, 422–448 (2019). https://doi.org/10.1016/j.ymssp.2018.09.019
8. Merino-Martínez, R., Sijtsma, P., Snellen, M., Ahlefeldt, T., Antoni, J., Bahr, C.J., Blacodon, D., Ernst, D., Finez, A., Funke, S., Geyer, T.F., Haxter, S., Herold, G., Huang, X., Humphreys, W.M., Leclère, Q., Malgoezar, A., Michel, U., Padois, T., Pereira, A., Picard, C., Sarradj, E., Siller, H., Simons, D.G., Spehr, C.: review of acoustic imaging methods using phased microphone arrays. CEAS Aeronaut. J. **10**(1), 197–230 (2019). https://doi.org/10.1007/s13272-019-00383-4
9. Quantitative Analysis of the German Drone Market. Drones in Germany: Market Study by Drone Industry Insights. Ed. by Verband unbemannte Luftfahrt. Berlin
10. Schroth, L., Wackwitz, K.: BVLOS Operations Report: Beyond Visual Line of Sight Report about Technology, Applications, Market Size, Unit Sales and Regulation. Drone Industry Insights, Hamburg
11. Martinetti, A., Schakel, E.J., van Dongen, L.A.M.: Flying asset. J. Qual. Main. Eng. **24**(2), 152–169 (2018). https://doi.org/10.1108/JQME-12-2016-0073. (**ISSN: 1355-2511**)
12. EASA. COMMISSION DELEGATED REGULATION (EU) 2019/945: on unmanned aircraft systems and on third-country operators of unmanned aircraft systems. 12 (March 2019)
13. EASA. COMMISSION IMPLEMENTING REGULATION (EU) 2019/947 - of 24 May 2019 - on the rules and procedures for the operation of unmanned aircraft: 2019/947
14. EASA. Annex I to ED Decision 2019/021/R
15. Olson, I., Atkins, E.M.: Qualitative failure analysis for a small quadrotor unmanned aircraft system. In: Guidance, Navigation,

and Control and Co-located Conferences. (2013). ISBN: 978-1-62410-224-0. https://doi.org/10.2514/6.2013-4761
16. Shraim, H., Awada, A., Youness, R.: A survey on quadrotors: configurations, modeling and identification, control, collision avoidance, fault diagnosis and tolerant control''. IEEE Aerosp. Electron. Syst. Mag. **33**(7), 14–335 (2018). https://doi.org/10.1109/MAES.2018.160246. (**ISSN: 0885-898**)
17. Fourlas, G.K., Karras, G.C.: A survey on fault diagnosis methods for UAVs. In: International Conference on Unmanned Aircarft Systems (ICUAS) (2021). https://doi.org/10.1109/ICUAS51884.2021.9476733
18. Puchalski, R., Giernacki., W (2022) UAV fault detection methods, state-of-the-art. Drones **6**(11), p. 330. https://doi.org/10.3390/drones6110330
19. Saied, M., Shraim, H., Francis, C.: A review on recent development of multirotor UAV fault-tolerant control systems. In: IEEE Aerospace and Electronic Systems Magazine (2024), pp. 1–30. ISSN: 0885-8985. https://doi.org/10.1109/MAES.2023.3327697
20. DJI. Matrice 200 Series V2 Maintenance Manual. Ed. by DJI. (2020). https://www.manualslib.com/manual/2018877/Dji-Matrice-300-Rtk.html#manual
21. Koschlik, A.-K., Meyer, H., Arts, E., Conen, P., Jacob, G. , Soria-Gomez, M., Kamtsiuris, A.A., Jilke, L., Aigner, J., Raddatz, F., Wende, G.: Towards an integrated vehicle health management for maintenance of unmanned air systems. In: 2023 International Conference on Unmanned Aircraft Systems (ICUAS). (2023)
22. Meissner, R., Meyer, H., Wicke, K: Concept and economic evaluation of prescriptive maintenance strategies for an automated condition monitoring system. In: (2021). https://api.semanticscholar.org/CorpusID:233475156
23. Soria-Gomez, M., Koschlik, A.-K., Arts, E., Raddatz, F., Wende, G.: Acoustic non-destructive testing of UAS's propellers during predeparture and post-flight checks. In: ECNDT 2023 (2023)
24. Kołodziejczak, M., Puchalski, R., Bondyra, A., Sladic, S., Giernacki, W.: "Toward lightweight acoustic fault detection and identification of UAV rotors". In: International Conference on Unmanned Aircarft Systems (ICUAS) (2023). https://doi.org/10.1109/ICUAS57906.2023.10156624
25. Bondyra, A., Kołodziejczak, M., Kulikowski, R., Giernacki, W.: An acoustic fault detection and isolation system for multirotor UAV. Energies **15**(11), 3955 (2022). https://doi.org/10.3390/en15113955
26. Altinors, A., Yol, F., Yaman, O.: A sound based method for fault detection with statistical feature extraction in UAV motors. Appl. Acoust. **183**, 108325 (2021). https://doi.org/10.1016/j.apacoust.2021.108325
27. Liu, W., Chen, Z., Zheng, M.: An audio-based fault diagnosis method for quadrotors using convolutional neural network and transfer learning. http://arxiv.org/pdf/2003.02649v2
28. Iannace, G., Ciaburro, G., Trematerra, A.: Fault diagnosis for UAV blades using artificial neural network. Robotics **8**(3), 59 (2019). https://doi.org/10.3390/robotics8030059
29. Nunes, E.C.: Anomalous sound detection with machine learning: a systematic review. http://arxiv.org/pdf/2102.07820v1
30. Yang, P., Wen, C., Geng, H., Liu, P.: Intelligent fault diagnosis method for blade damage of quad-rotor uav based on stacked pruning sparse denoising autoencoder and convolutional neural network. Machines (2021). https://doi.org/10.3390/machines9120360
31. CAE Software und Systems GmbH, ed. CAE Software and Systems. https://www.cae-systems.de/en/
32. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, Oriol, W., Pete, W., Martin, W., Martin, Y., Yuan, Z., Xiaoqiang : TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. (2015). https://www.tensorflow.org/
33. Dougherty, R.P.: Functional beamforming for aeroacoustic source distributions. In: 20th AIAA/CEAS Aeroacoustics Conference. Reston, Virginia: American Institute of Aeronautics and Astronautics, (2014). ISBN: 978-1-62410-285-1. https://doi.org/10.2514/6.2014-3066
34. Acoular Development Team. Acoular. http://acoular.org/install/index.html
35. Sijtsma, P.: Beamforming on moving sources. Brussels, (2007). https://reports.nlr.nl/server/api/core/bitstreams/e88b1b86-837c-4d37-bbf8-766ce0b51415/content
36. Chinchor, N.: MUC-4 evaluation metrics. In: Proceedings of the 4th Conference on Message understanding (1992). https://doi.org/10.3115/1072064.1072067