The Aeronautical Journal (2026), 1–20
doi:10.1017/aer.2026.10235

ROYAL AERONAUTICAL SOCIETY

RESEARCH ARTICLE

A low-cost wireless accelerometer approach to UAV propeller fault detection

Alican Yilmaz[1], Erkan Caner Ozkat[1] and Fatih Gul[2]

[1]Department of Mechanical Engineering, Faculty of Engineering & Architecture, Recep Tayyip Erdoğan University, Rize, Türkiye
[2]Department of Electrical and Electronics Engineering, Faculty of Engineering & Architecture, Recep Tayyip Erdoğan University, Rize, Türkiye
Corresponding author: Erkan Caner Ozkat; Email: erkancaner.ozkat@erdogan.edu.tr

Received: 30 May 2026 UTC; **Revised:** 24 July 2026 UTC; **Accepted:** 12 August 2026 UTC

Keywords: condition-based maintenance; machine learning; propeller fault detection; unmanned aerial vehicle; vibration monitoring

Abstract

Propeller failure in flight is one of the most safety-critical failure modes for small multirotor unmanned aerial vehicles (UAVs). It is sudden, generally undetectable on conventional telemetry and often catastrophic. This study addresses the problem through a simple onboard sensing scheme. A wireless triaxial accelerometer (ADXL345) mounted on the airframe streams vibration data to a ground station via an ESP32 microcontroller over Wi-Fi. Seven propeller conditions were investigated: one healthy reference and six deliberately introduced defects of varying severity. Each flight followed the same protocol consisting of hovering, controlled manoeuvring and harsh manoeuvring, with a nominal duration of 60 s per phase. After interquartile-range (IQR) based outlier removal, 27 statistical features were extracted from 10 s sliding windows (180 windowed samples in total, with two minority classes of fewer than 15 samples each) and supplied to four classical classifiers: random forest (RF), support vector machine (SVM), k-nearest neighbors (kNN) and decision tree (DT). RF achieved the best performance. For the binary healthy-versus-faulty task, RF reached 94.44% test accuracy and 96.67% mean accuracy under five-fold cross-validation. For the seven-class task, RF reached 79.63% test accuracy (91.67% CV). The binary result indicates the potential to support an in-flight return-to-launch decision rule after validation under broader flight conditions, whereas the seven-class task remains bottlenecked by dataset size rather than by classifier choice. The implications of this gap and the path towards onboard deployment on a Pixhawk-class flight controller are discussed.

Nomenclature

a	acceleration along a given axis, m/s^2
k	IQR multiplier used for outlier removal
M	vibration magnitude, $\sqrt{x_X^2 + x_Y^2 + x_Z^2}$
N	number of samples in a window
Q_1, Q_3	first and third quartiles of an axis
x_a	acceleration sample on axis a
ADXL345	analog devices three-axis MEMS digital accelerometer
ANN	artificial neural network
BVLOS	beyond visual line of sight
CBM	condition-based maintenance
CNN	convolutional neural network
CRISP-DM	cross-industry standard process for data mining

© The Author(s), 2026. Published by Cambridge University Press on behalf of Royal Aeronautical Society. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided that no alterations are made and the original article is properly cited. The written permission of Cambridge University Press or the rights holder(s) must be obtained prior to any commercial use and/or adaptation of the article.

Downloaded from https://www.cambridge.org/core. 18 Nov 2026 at 10:36:50, subject to the Cambridge Core terms of use.

CV	cross-validation
DT	decision tree
EASA	European Union Aviation Safety Agency
ESP32	low-power Wi-Fi/Bluetooth microcontroller
FAA	Federal Aviation Administration
HUMS	health and usage monitoring system
I^2C	inter-integrated circuit (serial bus)
IMU	inertial measurement unit
IQR	interquartile range
kNN	*k*-nearest neighbors
LSTM	long short-term memory
MEMS	micro-electro-mechanical system
P2P	peak-to-peak amplitude
RBF	radial basis function
RF	random forest
RMS	root mean square
SORA	specific operations risk assessment
SPI	serial peripheral interface
SVM	support vector machine
UAV	unmanned aerial vehicle

Greek symbol

μ	mean value of a windowed signal
σ	standard deviation of a windowed signal

1.0 Introduction

Small multirotor unmanned aerial vehicles (UAVs) have become an established component of civil aviation. They are used in power-line survey, crop monitoring, medical sample delivery, disaster-zone mapping and, in certain jurisdictions, last-mile parcel delivery. Enthusiasts no longer adoption; it is increasingly driven by operators who require a productive asset together with a corresponding maintenance plan [1]. The platform-level engineering choices that support these missions have themselves been the subject of focused study: aerodynamic optimisation of wing profiles for small fixed-wing UAVs [2], performance analysis of solar-augmented fixed-wing platforms under varying environmental conditions [3] and machine-learning-aided thrust [4] and autonomous-flight [5] optimisation for rotorcraft and helicopter platforms have all advanced in parallel.

The propeller is the component most exposed to that operating life. Blade tips contact branches and other obstacles, are subjected to bird strikes and may develop fatigue cracks after several hundred flights. A defect that is only inconvenient on a hobbyist quadrotor becomes a safety event on a 25 kg cargo platform operating over populated areas. The underlying failure dynamics are well documented: an imbalanced or chipped blade modifies the rotor first-order vibration signature long before structural shedding occurs, and that signature is detectable on the airframe [6, 7].

The conventional response is visual inspection between flights. This approach is slow, dependent on inspector expertise and provides no information about damage that may develop during the next sortie. As commercial operations move towards beyond visual line of sight (BVLOS) flight under the European Union Aviation Safety Agency (EASA) Specific Operations Risk Assessment (SORA) and the proposed Federal Aviation Administration (FAA) Part 108 framework, regulators have signalled that some form of in-flight health awareness is expected to be incorporated into the airworthiness case rather than delegated to ground crew [8, 9]. Condition-based maintenance (CBM) for UAV rotors is therefore not merely a research topic; it is moving towards becoming a regulatory requirement.

Vibration analysis is the natural sensing modality in this context. It is mature in helicopter health and usage monitoring systems (HUMS) and in wind turbine gearbox monitoring. The supporting hardware

Downloaded from https://www.cambridge.org/core. 18 Nov 2026 at 10:36:50, subject to the Cambridge Core terms of use.

has become sufficiently small and inexpensive to be carried by a 250 g airframe. A micro-electro-mechanical system (MEMS) triaxial accelerometer with a microcontroller front-end weighs only a few grams, and a serviceable wireless link can be assembled from off-the-shelf components at low cost. The pertinent engineering question is no longer whether vibration data can be streamed from a small UAV in flight, but how that data should be exploited once acquired.

To address this gap, the present study provides a focused answer to that question. A multirotor UAV was instrumented with an ADXL345 accelerometer and an ESP32 microcontroller streaming over Wi-Fi to a ground station. Seven configurations were flown: one healthy propeller and six deliberately damaged variants of increasing severity. Each flight followed the same 180 s protocol so that vibration records could be compared across conditions. Each condition is represented by a single flight, and the sliding-window feature extraction yields 180 samples in total. Four classical machine-learning classifiers were then applied to features extracted from 10 s sliding windows.

The methodological choice to remain within the classical regime is deliberate. Deep models are powerful but data-hungry; with 180 windowed samples in total, such models would overfit before generalising. Convolutional, recurrent and autoencoder architectures applied to vibration data routinely require several thousand training windows to converge reliably [10–12]. The objective here is to determine how far a simple and transparent pipeline can be pushed on the kind of dataset that an operator can realistically collect in a single afternoon.

Binary fault detection (healthy versus any defect) reaches 94.44% test accuracy with random forest (RF) and 96.67% mean accuracy under five-fold cross-validation. This level of performance indicates the potential to support an in-flight return-to-launch decision rule, although the false-positive and false-negative trade-off would need to be validated under broader flight conditions before deployment. The seven-class fault identification problem behaves differently: accuracy plateaus at 79.63% on the held-out test set, with most errors occurring among the low-amplitude faults (Fault-2, Fault-3, Fault-5) that share similar statistical fingerprints. The bottleneck in this case is the dataset rather than the algorithm.

The novelty of the present study lies in three concrete contributions. The first is the deployment of a low-cost wireless vibration sensing pipeline based on the ADXL345 accelerometer and an ESP32 microcontroller transmitting over Wi-Fi, which removes the need for tethered acquisition hardware during real flight. The second is a side-by-side evaluation of both binary (healthy versus any defect) and seven-class (fault-type identification) classification on the same dataset, which makes the resolution gap between the two tasks directly visible. The third is a demonstration that RF, applied to hand-crafted statistical features extracted from short sliding windows, remains a practical and interpretable choice under the limited-data regime that an individual operator can collect in a single test campaign.

The remainder of this paper is organised as follows. Section 2 reviews the relevant literature. Section 3 describes the hardware, the test protocol and the data pipeline. Section 4 presents the vibration analysis and the classifier results. Section 5 discusses the limitations and the deployment route. Section 6 concludes the study.

2.0 Related work

2.1 Vibration-based fault diagnosis in rotating machinery

Vibration analysis for rotating machinery is a mature field. The tutorial by Randall and Antoni on rolling-element bearing diagnostics remains a standard reference for the underlying signal processing [13]. The same principles, including envelope analysis, cyclostationary modelling and order tracking, transfer reasonably well to propeller analysis once a speed reference is available. Reviews by Liu et al. [14] and Lei et al. [15] document how the field has absorbed machine learning over the past decade. Sharma and Parey provide a more focused account of condition indicators in gear diagnosis [16]. A complementary strand of the literature concerns the wireless side of the problem, and the trade-offs typically identified for industrial wireless monitoring (battery life, packet loss and sampling-rate limits imposed by the radio link) extend almost directly to the small-UAV use case considered here.

Downloaded from https://www.cambridge.org/core. 18 Nov 2026 at 10:36:50, subject to the Cambridge Core terms of use.

Rangel-Magdaleno et al. employed discrete wavelet decomposition to detect partially damaged rotor bars in induction motors under varying load [17]. Although their setup is industrial rather than airborne, the central finding, that mid-band wavelet energies separate healthy from faulty conditions more cleanly than raw time-domain statistics, motivates the choice of statistical descriptors used in the present study. Simple moments were preferred over wavelet coefficients because the sampling rate available over the Wi-Fi link is too low to resolve high-frequency content.

2.2 Acoustic and IMU-based UAV propeller monitoring

Two complementary modalities have been investigated for small UAVs. Acoustic monitoring is attractive because it does not require an airframe-mounted sensor. Iannace et al. trained a neural network on microphone recordings of healthy and damaged quadrotor blades [18]. Altinors et al. later extended this approach to UAV motors using hand-crafted spectral features and reported high accuracy in a controlled acoustic environment [19]. The principal drawback is environmental: wind, ambient traffic and operator speech rapidly degrade microphone data in field conditions.

Inertial measurement unit (IMU) and accelerometer-based monitoring avoids this problem at the cost of additional payload. Bondyra et al. used the existing flight-controller IMU together with threshold-based rules to flag propeller damage [20]. Ghalamchi et al. developed a real-time vibration-based scheme for multicopters and reported convincing detection of imbalance [7]. Khalastchi et al. approached the broader anomaly-detection problem online using model-based residuals across heterogeneous sensors [21]. The work reported here belongs to the airframe-mounted tradition. It differs in that the sensor is a dedicated, externally clocked unit streaming through its own radio link, which decouples it from flight-controller bandwidth limits and from the firmware schedule of the autopilot.

2.3 Machine learning approaches for rotorcraft health monitoring

The machine-learning literature on rotorcraft health divides broadly into classical and deep approaches. On the classical side, ensemble trees, support vector machines (SVM), and k-nearest neighbors (kNN) have been used for decades in bearing and gear fault classification [22–25]. The original decision-tree (DT) formulation by Quinlan still underpins many of the interpretable pipelines used in industrial CBM [25]. On the deep side, Saufi et al. [10], Zhao et al. [11] and Khan and Yairi [12] survey the rapid adoption of convolutional neural networks (CNNs), long short-term memory (LSTM) networks and autoencoders for vibration data. Within aerospace specifically, recurrent architectures have been applied to mini-UAV navigation-phase flight-data estimation within a Cross-Industry Standard Process for Data Mining (CRISP-DM) workflow [26], hybrid artificial neural network (ANN) approaches combined with population-based optimisation have been used for emission tuning of gas turbines [27], and supervised regression has been used for aircraft takeoff-speed prediction from flight-parameter data [28]. A complementary observation from outside the aerospace domain is that systematic benchmarking of several classical classifiers on a single labelled dataset (as performed for earthquake-induced soil liquefaction classification using RF, SVM, kNN and gradient-boosted variants) consistently identifies ensemble trees as the most stable performer under modest sample size [29], which is one of the motivations for the model selection adopted in the present study.

For small UAVs specifically, Pourpanah et al. explored hybrid feature-selection schemes positioned between classical and deep regimes [30], and Wang et al. applied a hybrid prognostics scheme for bearings [31]. A recurring observation across these studies is that deep models reach strong accuracy when several thousand labelled windows are available, but degrade or fail to converge when the training set is on the order of one or two hundred windows. The classical pipeline considered here is positioned for the latter regime, which is the regime that an individual operator with a single test cell typically inhabits.

The present contribution differs from this body of work in several respects. The methodology deliberately remains within the classical regime because most recent UAV propeller-fault studies train deep models on data volumes that an operator cannot reasonably collect during commissioning, whereas the

Downloaded from https://www.cambridge.org/core. 18 Nov 2026 at 10:36:50, subject to the Cambridge Core terms of use.

Table 1. *Positioning of the present study against representative UAV propeller/rotor fault-detection work. Accuracies are not listed because they are not directly comparable across different platforms, defect sets and datasets*

Study	Sensing	Rate	Classes	Conditions
Iannace et al. [18]	Acoustic (microphone)	Audio	2	Controlled (lab)
Altinors et al. [19]	Acoustic (microphone)	Audio	Multi	Controlled acoustic
Ghalamchi et al. [7]	On-board IMU	High	2 (imbalance)	Indoor
Bondyra et al. [20]	On-board IMU, threshold	High	2	Laboratory
Present work	Dedicated MEMS, Wi-Fi link	∼10 Hz	2 and 7	Indoor, no wind

aim here is to characterise what a small operator can accomplish with one afternoon of test flights and a laptop. The flight protocol is structured around three consecutive phases (hovering, controlled manoeuvring and harsh manoeuvring), which allows phase-dependent behaviour to be quantified rather than averaged out. Both the binary and the multi-class versions of the classification problem are reported on the same dataset, which makes the performance gap between the two tasks directly visible.

To position the present study against representative prior UAV propeller and rotor fault-detection work, Table 1 summarises the sensing modality, the effective acquisition rate, the number of conditions and the test environment for each. A direct head-to-head accuracy comparison is deliberately avoided because each study uses a different platform, defect set, and dataset; the table is intended to clarify the methodological niche occupied by the present work rather than to rank accuracy. Within this landscape, the present study is distinguished by a dedicated, externally clocked accelerometer streaming over its own wireless link and by the joint reporting of binary and seven-class results on a single dataset.

3.0 Methodology

3.1 Hardware and test platform

The test platform is a small custom-built quadrotor with brushless DC motors and four fixed-pitch polymer propellers of 19 cm diameter. The take-off mass with sensor payload is approximately 1.2 kg. The motors are driven through electronic speed controllers, and a multi-cell lithium-polymer battery powers the platform. The autopilot is a standard hobby-grade flight controller operating in stabilised mode; it is not part of the data pipeline, and no autopilot logs are used in the present study.

The vibration sensor is an ADXL345 (analog devices), a three-axis MEMS digital accelerometer with selectable measurement range of ±2, ±4, ±8 or ±16 g and 13-bit resolution on each axis. For reference, the ±16 g full-scale option corresponds to approximately ±157 m/s^2; all accelerations in this paper are reported in m/s^2. It communicates with an ESP32 microcontroller over an inter-integrated circuit (I^{2}C) bus. The ESP32 streams the readings to a ground-station laptop over a Wi-Fi link. Figure 1 shows the transmitter unit; Figure 2 shows the healthy reference propeller with the sensor installed.

In principle the ADXL345 can sample at up to 3,200 Hz. In practice, with the Wi-Fi link included in the pipeline, the usable end-to-end sampling rate is approximately 10 Hz. This rate is a hard constraint of the wireless telemetry pipeline and precludes frequency-domain analysis of bearing or blade-pass harmonics, which lie well above 10 Hz on a multirotor of this scale. It does not, however, preclude the lower-band amplitude statistics adopted in the present study. The structural defect modes targeted here (static imbalance, leading-edge chips, asymmetric blade thinning) excite once-per-revolution and other low-order amplitude modulations that are aliased into the band below 10 Hz but remain visible as changes in the time-domain amplitude distribution: standard deviation, peak-to-peak (P2P) amplitude, root mean square (RMS) and the higher statistical moments are sensitive to these modulations even when the carrier harmonic itself is unresolved. Tables 4 and 9 confirm that this is the case in practice on the dataset considered, with healthy and high-amplitude fault classes already separable on the Y-axis amplitude statistics alone. It must be emphasised, however, that because the carrier harmonic is unresolved

Downloaded from https://www.cambridge.org/core. 18 Nov 2026 at 10:36:50, subject to the Cambridge Core terms of use.

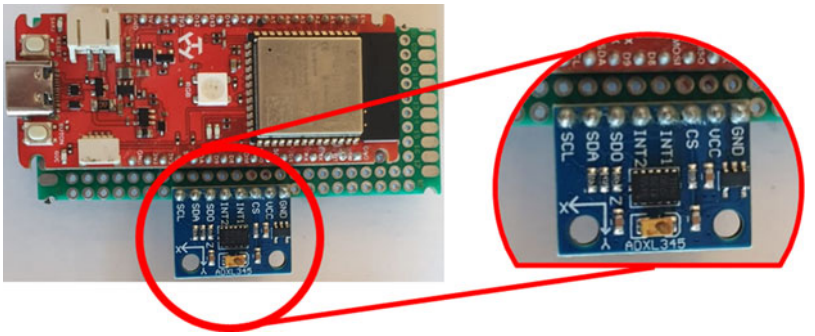


Figure 1. *ESP32-based wireless sensor transmitter with ADXL345 accelerometer, mounted on the UAV airframe close to the propeller attachment point.*



Figure 2. *Healthy reference propeller with the ADXL345 sensor and ESP32 transmitter installed.*

at this rate, the retained features capture aliased low-order amplitude modulations and may also incorporate low-frequency rigid-body airframe motion; the present results therefore establish detectability on the amplitude envelope rather than on resolved spectral lines, and the term 'vibration monitoring' should be read in this restricted sense. A short high-rate capture (onboard logging at the full ADXL345 bandwidth) for a healthy and a faulty case is planned to confirm that the 10 Hz amplitude statistics track the true broadband vibration energy.

The sensor block is mounted on the airframe close to one of the motor mounts, at an approximate distance of 4–5 cm from the propeller hub. This motor-mount position is a deliberate departure from the more common centre-of-gravity placement used in airframe-wide health monitoring. A centre-of-gravity sensor captures the structural propagation of vibrations averaged across all four rotors, which is appropriate for global airframe diagnostics. A motor-mount sensor, by contrast, preferentially captures the signature of the rotor closest to it, which is appropriate when the question of interest is whether one specific propeller is degraded. Since the present study targets single-propeller fault identification rather than global airframe health, the motor-mount position was preferred. The ESP32 is powered by

Downloaded from https://www.cambridge.org/core. 18 Nov 2026 at 10:36:50, subject to the Cambridge Core terms of use.

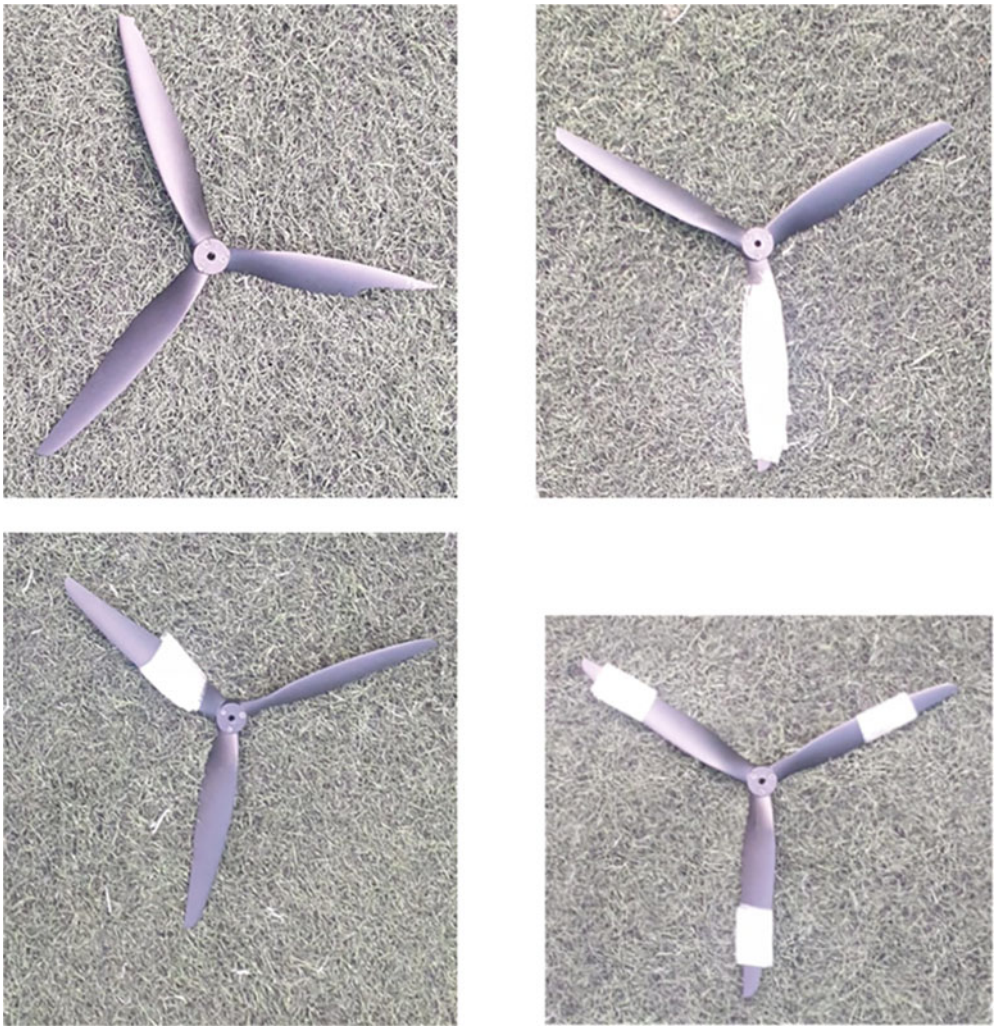


Figure 3. *Examples of the deliberately damaged propellers used in the campaign, illustrating the range of mass-perturbation and surface-modification defects introduced on the blades.*

an auxiliary USB power bank carried on the airframe, separate from the main flight battery, so that no current from the propulsion bus passes through the sensing electronics.

3.2 Propeller configurations

Seven propellers were tested: one healthy reference and six damaged propellers. The damage modes were introduced by applying controlled mass-perturbation patches at known positions along the propeller blades, which is a standard reversible surrogate technique for producing a calibrated imbalance that mimics in-service effects such as foreign-object adhesion, debris bonding and partial blade-surface modification. Representative photographs are shown in Figure 3. The modifications span single-blade tip-mass additions, asymmetric mid-blade perturbations and symmetric tip-mass additions distributed across all three blades; together they cover a range of imbalance configurations from highly asymmetric (causing strong once-per-revolution excitation) to nearly symmetric (causing primarily a global mass change with weak amplitude growth).

Downloaded from https://www.cambridge.org/core. 18 Nov 2026 at 10:36:50, subject to the Cambridge Core terms of use.

Table 2. *Mapping of fault identifiers to the vibration-amplitude class observed on the cleaned accelerometer record*

Fault ID	Damage mode	Amplitude class
Fault-1	Subtle blade-surface modification	High
Fault-2	Single-blade tip mass addition	Low
Fault-3	Localised mass perturbation	Low
Fault-4	Asymmetric mid-blade mass addition	High
Fault-5	Symmetric tip-mass addition on all three blades	Low
Fault-6	Severe asymmetric mass perturbation	High

Defects are referred to throughout as Fault-1 to Fault-6. Table 2 maps each fault identifier to its damage mode and to the empirical vibration-amplitude class observed in the data (Tables 3 and 4).

A useful first observation, available before any classifier is trained, is that the six defects fall into two empirically distinct groups in the raw vibration record (see Tables 3 and 4; Figures 4 and 5). Fault-1, Fault-4 and Fault-6 produce high-amplitude, broad-distribution vibration signatures consistent with strongly asymmetric mass perturbations. Fault-2, Fault-3 and Fault-5 produce low-amplitude signatures whose statistics overlap with the healthy reference, consistent with localised or nearly symmetric perturbations that do not strongly excite the once-per-revolution component. This severity-based grouping is referred to in the discussion of the multi-class confusion matrix, where most of the residual error of the seven-class classifier occurs within the low-amplitude group.

3.3 Flight protocol

Each flight followed the same 180 s protocol. The first 60 s consists of steady hovering at a fixed altitude. The middle 60 s consists of controlled manoeuvring with gentle yaw, pitch, and translation inputs. The final 60 s consists of harsh flight, with deliberately aggressive stick inputs and rapid reversals. The same test pilot performed every run. The test cell is a netted indoor volume with no wind, which removes a major source of confounding low-frequency vibration.

The three phases were selected because they place qualitatively different mechanical demands on the rotor. Hovering is closest to a textbook steady state and provides the cleanest fault signature. Manoeuvring introduces controlled changes in rotor rotational speed and thrust vector. Harsh flight stresses the blade with rapid commanded accelerations and recovers some of the worst-case vibration encountered in real operational use.

Two of the runs produced shorter usable records than the nominal 180 s protocol: 42.86 s for Fault-2 and 61.44 s for Fault-3. The partial data were retained for analysis.

3.4 Preprocessing

The raw stream contains two artefacts that need to be removed before feature extraction: a sensor warm-up transient at the start of each file, and occasional spurious outliers from telemetry hiccups. An interquartile-range (IQR) rule is applied on each axis independently.

For each axis $a \in \{X, Y, Z\}$,

$$\mathrm{IQR}_a = Q_{3,a} - Q_{1,a}, \tag{1}$$

and a sample x_a is retained only when

$$Q_{1,a} - k\,\mathrm{IQR}_a \le x_a \le Q_{3,a} + k\,\mathrm{IQR}_a, \tag{2}$$

for all three axes simultaneously. The multiplier was set to $k = 2.0$. With $k = 1.5$ (Tukey's default) an excessive proportion of high-amplitude harsh-flight samples were discarded. With $k = 3.0$ obvious spike

Downloaded from https://www.cambridge.org/core. 18 Nov 2026 at 10:36:50, subject to the Cambridge Core terms of use.

Table 3. *Data preprocessing summary. Sample counts per propeller before and after IQR-based outlier removal*

Propeller	Original	After cleaning	Removed (%)
Healthy	1846	1677	9.15
Fault-1	1850	1681	9.13
Fault-2	415	354	14.70
Fault-3	592	394	33.45
Fault-4	1840	1692	8.04
Fault-5	1736	1685	2.94
Fault-6	1283	1182	7.87

artefacts were preserved. The value of 2.0 represents a compromise and could be tuned per flight, but it was kept fixed across the dataset for reproducibility. It should be noted that the largest removal fraction occurs for Fault-3 (33.45%), one of the low-amplitude classes; the IQR rule may therefore attenuate some genuine harsh-flight transients, and the residual misclassification of the low-amplitude group should be read with this in mind. A sensitivity analysis of the final classification accuracies to $k \in \{1.5, 2.0, 3.0\}$ is planned for the extended dataset to confirm that the reported results are not an artefact of the chosen multiplier.

After outlier removal, each flight was clipped to the 0–180 s window so that all conditions cover the same nominal interval. Table 3 summarises sample counts before and after cleaning. The removal fraction is small (below 10%) for most flights and large (33%) for Fault-3, which had the noisiest stream.

The accelerometer signals were not explicitly gravity-compensated. The static gravity vector and any fixed mounting-angle offset appear only in the per-axis mean; the feature-importance analysis (section 4.3) shows that the three axis means carry negligible discriminative weight (each below 0.008), and the amplitude-dispersion features on which the classifiers rely (standard deviation, RMS, peak-to-peak amplitude) are by construction invariant to a constant offset. Because the sensor was mounted once and its orientation was held fixed across every propeller condition, any mounting-angle difference is common-mode and does not confound the between-condition comparison; the sensor was not removed or repositioned between runs.

3.5 Statistical characterisation

Table 4 reports summary statistics for the Y-axis, which is the axis carrying the strongest fault signature in the present configuration. The healthy propeller exhibits the lowest standard deviation, approximately 3.55 m/s^2. Fault-1, Fault-4 and Fault-6 are markedly different, with standard deviations between 17 and 26 m/s^2 and peak-to-peak amplitudes exceeding 100 m/s^2. Fault-2, Fault-3 and Fault-5 sit much closer to the healthy condition in raw amplitude. However, Fault-2 and Fault-3 show positive kurtosis (1.91 and 2.76), indicating heavier-than-Gaussian tails. This heavy-tailed character is the foothold that a classifier requires for these subtler defects.

3.6 Feature extraction

Features were computed on 10 s sliding windows with 5 s overlap. The 10 s length was selected as a compromise between three competing requirements: the window must contain enough samples (approximately 100 at the effective 10 Hz rate) for the higher statistical moments to stabilise; it must be short enough that no more than one of the three flight phases (hovering, controlled manoeuvring, harsh manoeuvring) is captured per window, which would otherwise blur the phase-dependent signature; and it must permit a sufficient number of windows per flight to support training, given that the nominal flight duration is 180 s and that two of the faulty-propeller runs produced shorter usable records. The 5 s overlap (50%) doubles the number of windows per flight; the resulting correlation between temporally

Downloaded from https://www.cambridge.org/core. 18 Nov 2026 at 10:36:50, subject to the Cambridge Core terms of use.

Table 4. *Statistical summary of cleaned accelerometer data on the Y-axis. Units: m/s^2, except kurtosis (dimensionless)*

Propeller	Mean	Std	P2P	Kurtosis
Healthy	−1.64	3.55	20.24	−0.21
Fault-1	−1.35	17.17	134.19	−0.50
Fault-2	−2.75	2.52	19.06	1.91
Fault-3	−1.66	2.85	18.75	2.76
Fault-4	−1.10	25.69	109.05	−0.96
Fault-5	−1.42	2.64	17.42	−0.16
Fault-6	−0.41	25.64	139.65	−0.16

Table 5. *Feature-vector class distribution after windowing*

Class	Samples
Healthy	34
Fault-1	34
Fault-4	34
Fault-5	34
Fault-6	26
Fault-3	11
Fault-2	7
Total	**180**

adjacent feature vectors and its consequence for the train/test split is addressed in section 5.6. A formal sensitivity sweep of the window length (for example, 5 s, 10 s and 15 s) is left for future work. For each axis, eight statistics were computed: mean, standard deviation, RMS, P2P amplitude, skewness, kurtosis, minimum and maximum. The magnitude signal $M = \sqrt{x_X^2 + x_Y^2 + x_Z^2}$ contributes three further descriptors (mean, standard deviation and RMS). This magnitude is formed from the per-axis samples after IQR outlier removal, so its descriptors inherit the same cleaning as the per-axis features. This yields $8 \times 3 + 3 = 27$ features per window.

Across all seven propellers, the sliding window produces 180 feature vectors in total. The class distribution is uneven (Table 5). The four longest-running conditions – Healthy, Fault-1, Fault-4 and Fault-5 – each contribute 34 windows. Fault-6 contributes 26. Fault-3 and Fault-2 contribute only 11 and 7 windows, respectively, owing to their shorter usable records. The implications of this imbalance are revisited in the discussion.

Four classical machine learning classifiers were evaluated: RF [22], SVM [23], kNN [24] and DT [25]. The RF model used 100 trees with bootstrap sampling and feature subsampling at each split, and final predictions were obtained by majority voting. The SVM model used a radial basis function (RBF) kernel. The kNN classifier was applied with $k = 5$ using the standardised feature space. The DT classifier used information-gain-based splitting.

Before training, all features were standardised using z-score normalisation. To avoid data leakage, the scaling parameters were fitted only on the training partition and then applied to the test partition. For cross-validation, standardisation was refitted separately within each training fold. The dataset was divided into training and test sets using a stratified 70/30 split. In addition, five-fold stratified cross-validation (CV) was performed on the full dataset to check whether the held-out test result depended on a particular random split. Since each propeller condition corresponds to a single continuous flight, neither the random split nor the stratified folds enforce flight-level grouping. The implications of this limitation are discussed explicitly in section 5.6.

Downloaded from https://www.cambridge.org/core. 18 Nov 2026 at 10:36:50, subject to the Cambridge Core terms of use.

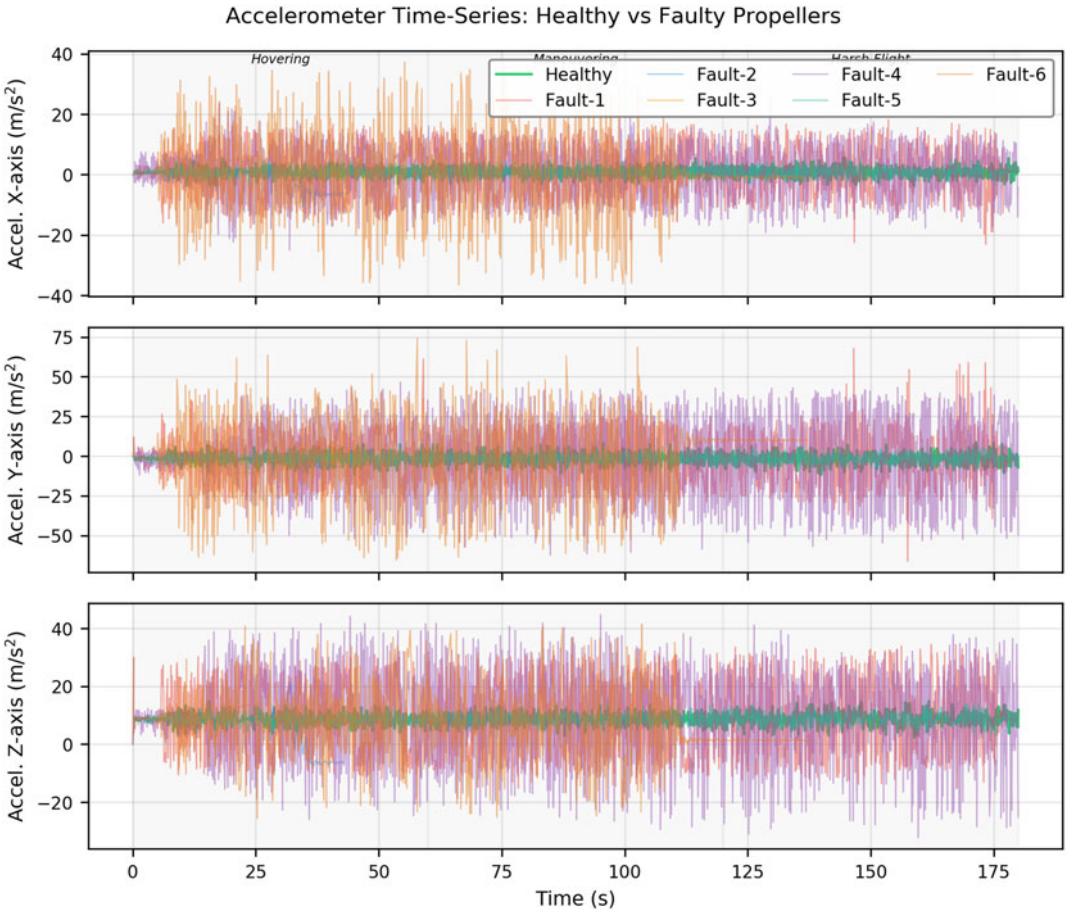


Figure 4. *Time-series accelerometer data comparing healthy and faulty propellers across the three flight phases (hovering 0–60 s, manoeuvring 60–120 s, harsh flight 120–180 s).*

4.0 Results

Throughout this section, classifier accuracies are reported in the running text as percentages to two decimal places and in the tables as fractional scores to four decimal places; the cross-validation results in Table 9 are given as mean ± standard deviation to three decimal places. Precision, recall and F1-score are computed per class and then aggregated in two ways: the *macro* average is the unweighted arithmetic mean of the per-class scores, which weights every class equally irrespective of its support, whereas the *weighted* average scales each per-class score by that class's support before averaging, so that the abundant high-amplitude classes dominate. Micro-averaging is not reported separately because, in the single-label multi-class setting used here, the micro-averaged precision, recall and F1-score all coincide with the overall accuracy. The two averages are reported together because they answer different questions on this imbalanced set.

4.1 Time-domain behaviour

Figure 4 shows representative time series for all conditions across the three flight phases. The healthy trace is the quietest. Fault-1, Fault-4 and Fault-6 visibly diverge from it within the first few seconds of hovering. Fault-2, Fault-3 and Fault-5 are not apparent by visual inspection of the raw trace, although subtle differences emerge once moments are computed.

Downloaded from https://www.cambridge.org/core. 18 Nov 2026 at 10:36:50, subject to the Cambridge Core terms of use.

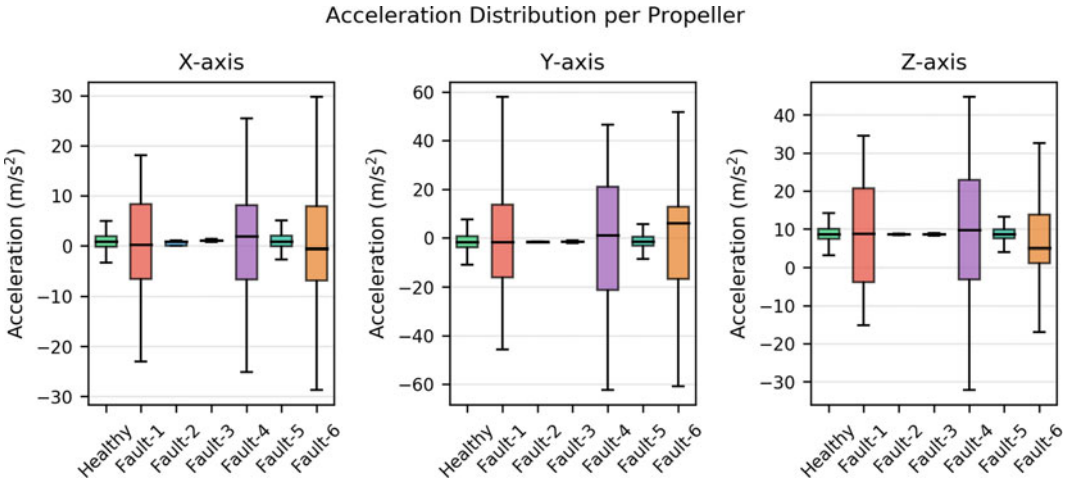


Figure 5. *Box-plot distributions of acceleration per propeller for the X, Y and Z axes after outlier removal. The box marks the interquartile range with the median line, the whiskers extend to 1.5 times the IQR, and points beyond are shown individually.*

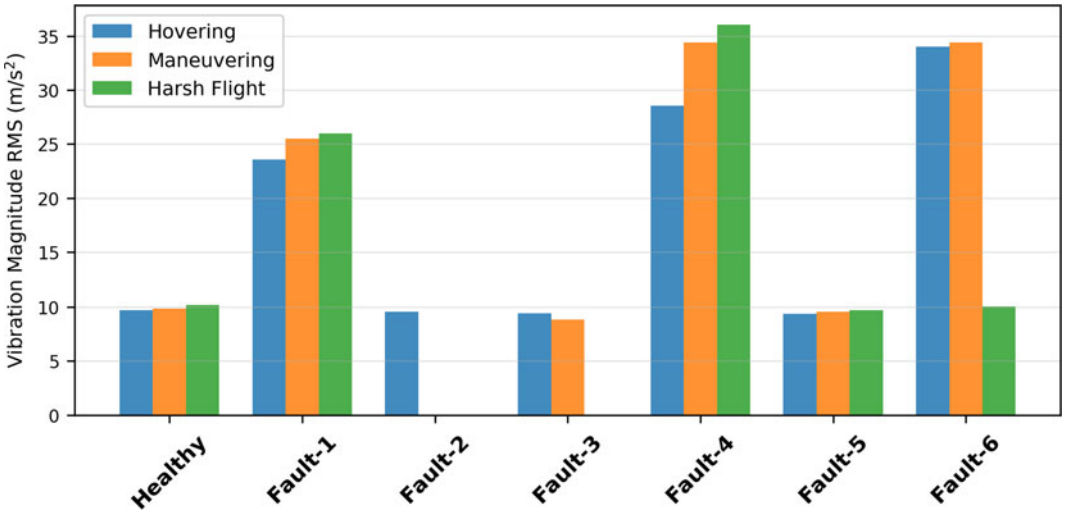


Figure 6. *Vibration magnitude (RMS) across flight phases for each propeller condition.*

Figure 5 shows the per-axis distributions after outlier removal. The Y-axis carries the strongest signal across most defects. This is consistent with the sensor mounting orientation, which aligns the Y-axis approximately with the dominant in-plane oscillation direction at the motor mount.

Figure 6 compares vibration magnitude RMS across the three flight phases. As expected, the harsh-flight phase is louder than hovering for every propeller. Notably, the contrast between healthy and faulty is largest in the hovering phase. In this phase the airframe is closest to a steady state and the fault-induced signal is not masked by manoeuvring-induced dynamics. For an in-flight monitoring scheme this is a favourable observation: the cleanest signal is available precisely during the steady periods, which is also when a control system can most readily take action.

4.2 Multi-class classification

Table 6 reports the seven-class results on the held-out test set. RF reaches 79.63% accuracy and an F1 of 77.87%. DT matches the same accuracy with a marginally higher F1 of 78.63%, while kNN and

Downloaded from https://www.cambridge.org/core. 18 Nov 2026 at 10:36:50, subject to the Cambridge Core terms of use.

Table 6. *Multi-class classification results (7 classes) on the held-out 30% test set*

Classifier	Acc.	Prec.	Recall	F1
RF	**0.7963**	0.7886	**0.7963**	0.7787
DT	0.7963	**0.7959**	0.7963	**0.7863**
kNN	0.7593	0.7766	0.7593	0.7357
SVM	0.7037	0.6575	0.7037	0.6676

Note: Bold values indicate the best result in each column.

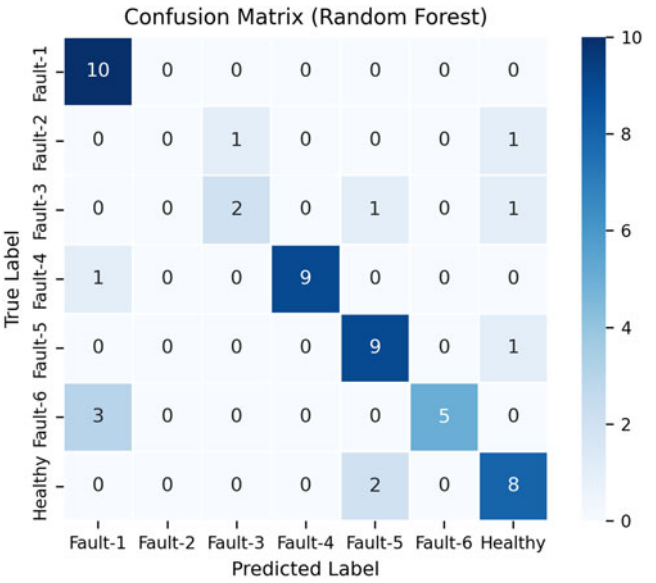


Figure 7. *Confusion matrix for the RF classifier on the seven-class task.*

SVM trail at 75.93% and 70.37%, respectively. The confusion matrix in Figure 7 explains the origin of the errors. They concentrate among Fault-2, Fault-3 and Fault-5, the three low-amplitude faults whose univariate moments are similar to the healthy baseline. The high-amplitude faults (Fault-1, Fault-4, Fault-6) are almost never confused with each other or with the healthy class. On this seven-class task, the difference between RF and DT lies within the resolution of a 54-sample test set: the two share an identical accuracy of 79.63%, and DT is in fact marginally ahead on F1. They should therefore be regarded as statistically comparable here, and the clear separation of RF from the field is established on the binary task and under cross-validation (Table 9) rather than on this single multi-class split. A formal pairwise comparison (for example McNemar's test, with bootstrap confidence intervals on the test metrics) is planned once additional flights enlarge the test set.

To support a class-level reading of the confusion matrix, Table 7 reports the per-class precision, recall and F1-score of the RF classifier on the seven-class task.

The class-wise breakdown confirms the pattern already visible in the confusion matrix. The Healthy class and the three high-amplitude faults (Fault-1, Fault-4, Fault-6) are recovered with strong precision and recall, with Fault-4 reaching perfect precision at 90% recall and Fault-1 reaching perfect recall at 71% precision. The three low-amplitude faults drive almost all of the residual error of the seven-class classifier: Fault-3 is recovered at 67% precision and 50% recall, and Fault-2 is not recovered at all (precision and recall both zero on the test split). The Fault-2 collapse is principally a sample-size artefact: only seven windows of Fault-2 exist in the entire dataset, of which two end up in the test set, and an ensemble of DTs trained on the remaining five minority-class samples produces no decision path that reaches Fault-2 with majority vote at inference time. Macro-averaged F1 of 67.16%

Downloaded from https://www.cambridge.org/core. 18 Nov 2026 at 10:36:50, subject to the Cambridge Core terms of use.

Table 7. *Per-class precision, recall and F1-score of the RF classifier on the 7-class held-out test set (54 trials in total)*

Class	Precision	Recall	F1-score	Support
Healthy	0.7273	0.8000	0.7619	10
Fault-1	0.7143	1.0000	0.8333	10
Fault-2	0.0000	0.0000	0.0000	2
Fault-3	0.6667	0.5000	0.5714	4
Fault-4	1.0000	0.9000	0.9474	10
Fault-5	0.7500	0.9000	0.8182	10
Fault-6	1.0000	0.6250	0.7692	8
Macro avg	0.6940	0.6750	0.6716	54
Weighted avg	0.7886	0.7963	0.7787	54

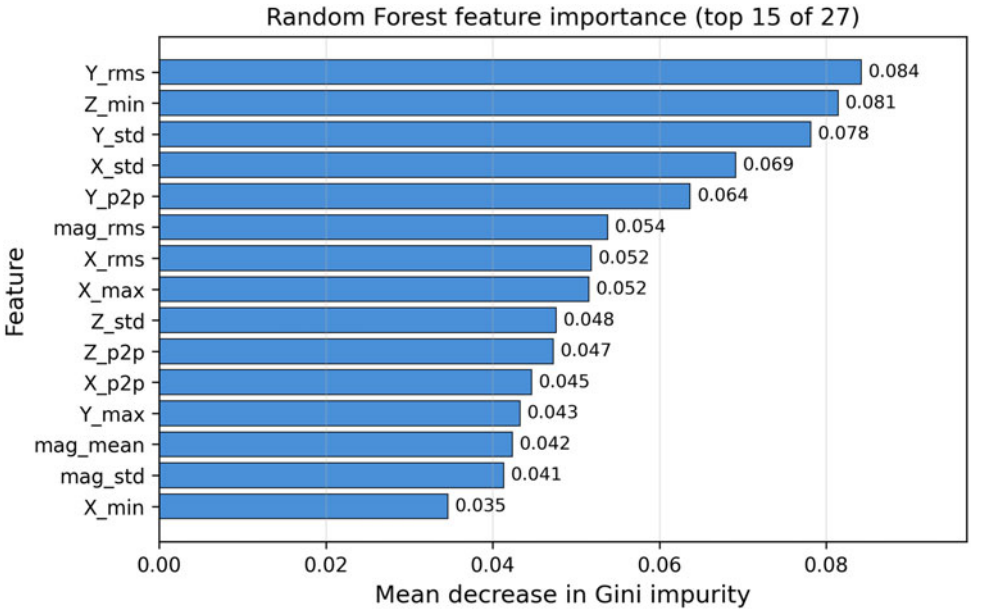


Figure 8. *RF Gini-based feature importance on the seven-class task, top 15 of 27 features. Amplitude-dispersion descriptors (RMS, standard deviation, P2P amplitude) dominate; the means of the three axes carry negligible weight.*

(versus weighted F1 of 77.87%) therefore overstates the difficulty of the high-amplitude classes and understates the difficulty of the minority classes; both averages should be read together.

4.3 Feature importance

RF provides an inexpensive estimate of feature relevance through the mean decrease in Gini impurity computed across the ensemble. Figure 8 reports the ranking on the seven-class task. The dominant predictor is the Y-axis RMS (0.084), closely followed by the Z-axis minimum (0.081), Y-axis standard deviation (0.078), X-axis standard deviation (0.069) and Y-axis P2P amplitude (0.064). The vibration-magnitude RMS (0.054) and X-axis RMS (0.052) appear in the next tier together with the X-axis maximum and the Z- and X-axis P2P amplitudes. The higher moments (skewness and kurtosis) sit consistently in the lower half of the ranking, with no single moment exceeding 0.020. The means of the

Downloaded from https://www.cambridge.org/core. 18 Nov 2026 at 10:36:50, subject to the Cambridge Core terms of use.

Table 8. *Binary classification results (healthy vs. any fault) on the held-out 30% test set*

Classifier	Acc.	Prec.	Recall	F1
RF	**0.9444**	**0.9432**	**0.9444**	**0.9433**
DT	0.9074	0.9116	0.9074	0.9091
kNN	0.8704	0.8658	0.8704	0.8677
SVM	0.8148	0.6639	0.8148	0.7317

Note: Bold values indicate the best result in each column.

Table 9. *Five-fold stratified cross-validation accuracy. Mean ± standard deviation across folds*

Classifier	Multi-class CV	Binary CV
RF	**0.917 ± 0.030**	**0.967 ± 0.021**
DT	0.883 ± 0.041	0.900 ± 0.048
kNN	0.783 ± 0.044	0.861 ± 0.047
SVM	0.778 ± 0.050	0.811 ± 0.011

Note: Bold values indicate the best result in each column.

three axes carry almost no discriminative weight (each below 0.008), which is expected given that the per-trial outlier removal and the per-window detrending implicit in the standard deviation calculation already absorb the mean offset. The combined evidence of Tables 4 and 7 together with this ranking supports the use of an amplitude-dispersion-based feature set rather than a centre-of-distribution feature set for this monitoring task.

4.4 Binary classification

Table 8 reports the healthy-versus-any-fault task. Every classifier improves noticeably. RF reaches 94.44% accuracy with an F1 of 94.33%. DT reaches 90.74%. kNN and SVM reach 87.04% and 81.48%, respectively. The improvement from multi-class to binary is between 11 and 15 percentage points across classifiers. The information required to indicate that 'something is wrong' is much more readily extracted from the data than the information required to identify which specific defect is present.

4.5 Cross-validation

Held-out scores from a single split can be misleading on a small dataset. Table 9 reports the five-fold stratified cross-validation results. RF yields a mean accuracy of 91.67% on the multi-class task and 96.67% on the binary task. The standard deviations are sufficiently small that the ranking among classifiers remains stable across folds. Figure 9 visualises the same information.

Figure 10 places the held-out and CV results side by side. The ensemble methods occupy the top of the ranking in both views. SVM is the weakest, which is attributed to the small dataset combined with the 27-dimensional feature space; RBF SVMs require careful tuning of C and γ that a 180-sample dataset cannot adequately support. kNN sits in the middle and is penalised by the high-feature dimensionality relative to sample count.

5.0 Discussion

5.1 Why RF performs best here

RF behaves well on the type of data considered. The features are heterogeneous – means and higher moments lie on quite different numerical scales – the sample size is small, and the class boundary is nonlinear. Trees handle heterogeneous scales without preprocessing tricks. Bagging across 100 trees yields a usable variance reduction even with fewer than 200 windows. In addition, the Gini importance

Downloaded from https://www.cambridge.org/core. 18 Nov 2026 at 10:36:50, subject to the Cambridge Core terms of use.

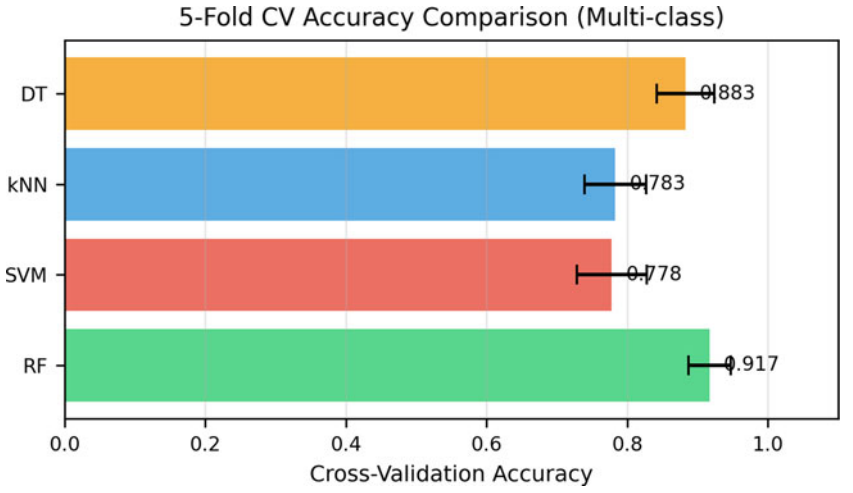


Figure 9. *Five-fold stratified cross-validation results for all classifiers on both classification tasks.*

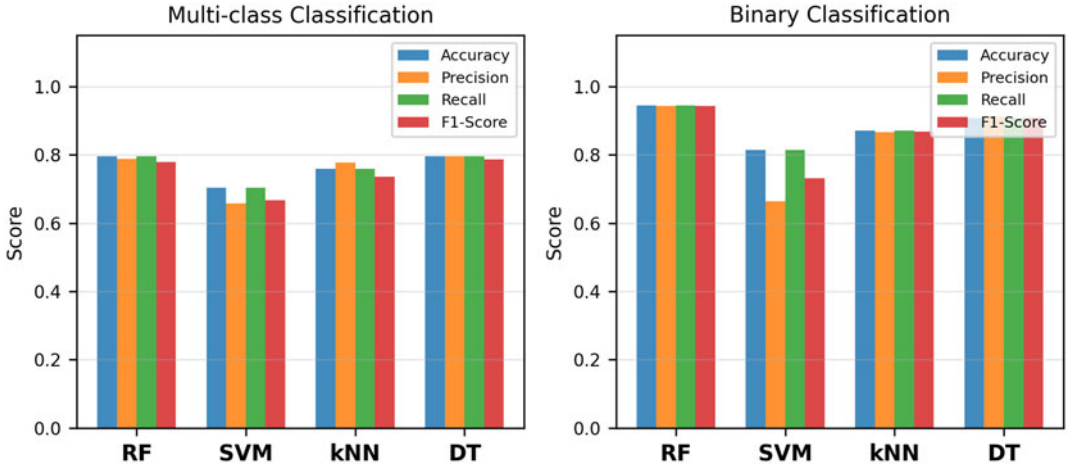


Figure 10. *Side-by-side comparison of classifier performance on the multi-class and binary tasks.*

provides a free interpretability benefit that supports the qualitative analysis in Section 4. This combination of an explicit feature definition followed by a compact classical model has been adopted productively in other analytical-to-AI pipelines outside the UAV domain. A recent example is the entropy-generation analysis of pulsating non-Newtonian heat flow [32], where the analytical model defines the descriptors and a learned model selects between regimes.

DT performs almost as well on the binary task but is more brittle in the multi-class setting. SVM and kNN both suffer from the dimensionality-to-sample-size ratio. None of these observations is surprising; the more interesting question concerns the magnitude of the gap between the binary and multi-class tasks.

5.2 The binary–multi-class gap

On the binary task, the classifier is effectively asked to separate a low-vibration cluster (Healthy, together with the low-amplitude Fault-2/3/5 group when quiescent) from a high-vibration cluster (Fault-1/4/6).

Downloaded from https://www.cambridge.org/core. 18 Nov 2026 at 10:36:50, subject to the Cambridge Core terms of use.

This separation is robust, and even simple statistics such as the Y-axis standard deviation are sufficient to recover most of it. On the multi-class task, the classifier must discriminate within those clusters, and within the low-vibration cluster the univariate moments are insufficient.

The multi-class accuracy could potentially be improved by adopting a more expressive feature set: frequency-domain features (if the sampling rate permitted), envelope spectra or learned representations. However, the binding constraint is judged to be data volume rather than feature design. With only seven windows for Fault-2, the classifier has very little to learn from. A useful next experiment would be to fly several short repeat runs of the low-amplitude faults rather than a single long run for each.

5.3 Rationale for not using SMOTE

Class imbalance is non-trivial here: 34 windows for Healthy against 7 for Fault-2. The standard response is to oversample with SMOTE or a related technique. This option was not adopted, for two reasons. First, the **synthetic minority oversampling technique** (SMOTE) synthesises new samples by interpolating between existing minority-class neighbours in feature space, but with only seven Fault-2 windows the neighbour structure is too sparse to support meaningful interpolation without introducing artefacts. Second, the class imbalance reflects a real physical asymmetry: severe faults cut flights short, and the operational use case of the classifier already accommodates the possibility of a flight ending early. Cost-sensitive learning, or stratified collection of additional data, are cleaner remedies than synthetic oversampling.

5.4 Comparison with deep learning on similar platforms

A natural question concerns how the present approach compares with deep methods on UAV vibration data. Reviews of deep learning for machinery health monitoring report that convolutional and recurrent architectures generally need training sets one to two orders of magnitude larger than the one used here before they begin to outperform careful classical baselines [10–12]. A parallel observation is made in the recent aerospace prognostics literature: deep and similarity-based pipelines for turbofan remaining-useful-life prediction on the Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) benchmark improve over classical baselines only once the training set covers a substantial portion of the degradation trajectory [33, 34], and the benefit is much weaker when only a handful of run-to-failure traces are available. The pattern is consistent: deep models scale well when data are plentiful and become wasteful or unstable when they are not. For a small operator able to conduct an afternoon of test sorties, the present result suggests starting with RF on hand-crafted moments and moving to a deep model only when the data budget justifies it. It should be acknowledged that no deep baseline was trained on the present dataset; the comparison above rests on the cited literature rather than on a controlled experiment on these 180 windows, and a compact one-dimensional convolutional baseline on the same protocol is planned so that the overfitting argument can be made empirically. Relatedly, because the Gini ranking (section 4.3) concentrates the discriminative weight in roughly five amplitude-dispersion features, a reduced or explicitly selected feature set is expected to particularly benefit the distance- and margin-based classifiers (kNN and SVM) that are penalised by the present 27-dimensional space; a feature-selection study is left for the extended dataset.

5.5 Towards onboard real-time deployment

The classifier itself is computationally inexpensive. A trained RF with 100 shallow trees fits readily on a Pixhawk-class or Cube Orange-class flight controller, and the 27 features can be computed on a rolling 10 s buffer in tens of milliseconds. The more demanding engineering questions lie upstream: a higher-rate accelerometer link without the present Wi-Fi packet-loss issue, a wired serial peripheral interface (SPI) or I^2C connection from the sensor into the autopilot, and a clear policy on the autopilot response when the classifier raises a flag. A plausible candidate policy is a ‘return to launch’ instruction

Downloaded from https://www.cambridge.org/core. 18 Nov 2026 at 10:36:50, subject to the Cambridge Core terms of use.

triggered by persistent binary fault detection across several consecutive windows, which keeps the false-alarm cost low and leaves the operator with an opportunity to override. The validity of such a policy depends on the false-positive and false-negative rates measured under conditions broader than the present indoor campaign, and must be established empirically before any deployment. In this sense the wireless pipeline demonstrated here should be read as a low-cost demonstrator that establishes the amplitude-based detectability floor, rather than as the production sensing architecture; a fielded monitor is expected to adopt the wired, higher-rate link discussed above. For a safety function the governing metric is the false-negative (missed-fault) rate rather than the overall accuracy, and this rate together with its confidence interval must be characterised under broader conditions before any return-to-launch rule is enabled. No safety claim is made on the basis of the present indoor dataset.

The transferability of the present pipeline to fixed-wing UAVs is an open question. A single-propeller fixed-wing platform reframes the multi-class problem from 'which rotor is degraded' to 'which damage mode is present' on the one propulsor. The vibration channel may also be less crowded because rotor-to-rotor harmonic interactions are not present. Whether these factors translate into improved detection or classification accuracy in practice is not something the present multirotor dataset can establish; targeted experiments on a fixed-wing platform would be required to address the question directly.

5.6 Limitations

Several limitations should be noted directly. The dataset is small, comprising 180 windows in total, of which two classes contain fewer than 15 samples each. Each propeller condition was also represented by a single flight rather than by repeated flights under the same condition, so the present dataset does not separate within-condition flight-to-flight variability from between-condition fault signature variability. Future campaigns should therefore acquire multiple independent flights per fault condition, in addition to longer individual recordings, in order to support a proper repeated-measures analysis of classifier confidence. The end-to-end sampling rate of approximately 10 Hz imposed by the Wi-Fi link precludes any frequency-domain analysis above the lower harmonics of the rotor. The test conditions are benign, consisting of indoor flights with no wind and a single pilot; outdoor operation in gusty conditions would introduce a low-frequency confounder that the present preprocessing does not address. Three further limitations follow from the single-flight design and must temper the interpretation of the reported accuracies. First, because the windows are drawn from one continuous flight per condition with 50% overlap, the random 70/30 split and the stratified cross-validation may place temporally adjacent, and therefore correlated, windows in both the training and the test partitions; the reported accuracies should accordingly be interpreted as a measure of within-flight discriminability, and a group-aware evaluation (leave-one-flight-out, with several independent flights per condition) is required before any claim of between-flight operational generalisation can be made. Second, the single motor-mount accelerometer observes primarily the rotor nearest to it, so the present results characterise detection when the degraded propeller is the instrumented one; detecting damage on a non-instrumented rotor would require additional sensors or the exploitation of cross-airframe propagation. Third, the defects are reversible mass-perturbation surrogates, which reproduce imbalance-driven excitation but not the stiffness and aeroelastic changes associated with genuine material loss such as cracks or leading-edge erosion; at least one real material-removal defect will be included in the extended campaign to anchor the surrogate against an in-service fault. The next iteration of the experimental campaign is planned to address these limitations together.

6.0 Conclusion

A small, wireless vibration-monitoring system was developed for a multirotor UAV and used to classify seven propeller conditions with classical machine learning. RF, applied to 27 statistical features per 10 s window, achieved 94.44% binary accuracy (96.67% under five-fold CV) and 79.63% seven-class accuracy (91.67% under CV). The binary result is high enough to indicate the potential of

Downloaded from https://www.cambridge.org/core. 18 Nov 2026 at 10:36:50, subject to the Cambridge Core terms of use.

a persistent-detection return-to-launch rule, subject to validation under broader flight conditions. The multi-class result is limited principally by the size of the dataset.

Hovering emerges as the most informative phase for fault detection, which aligns well with the steady-state requirements of an in-flight monitor. The substantial gap between binary and multi-class performance indicates that the appropriate path to improvement is to collect more data per defect rather than to substitute a more complex model. Future work will target fixed-wing platforms, wired high-rate sensing into a Pixhawk-class autopilot and outdoor flight in non-benign conditions.

Data availability statement. The accelerometer recordings and the derived feature dataset generated during this study are available from the corresponding author on reasonable request.

Acknowledgements. The authors would like to express their deepest gratitude to the Recep Tayyip Erdoğan University Development Foundation and the Recep Tayyip Erdoğan University Scientific Research Projects Coordinator Unit (BAP) for their generous financial support and steadfast institutional commitment.

Author contributions. Conceptualisation, A.Y. and E.C.O.; methodology, A.Y. and E.C.O.; software, A.Y., E.C.O. and F.G.; validation, A.Y., E.C.O. and F.G.; formal analysis, A.Y. and E.C.O.; investigation, A.Y. and E.C.O.; resources, A.Y., E.C.O. and F.G.; data curation, A.Y., E.C.O. and F.G.; writing—original draft preparation, A.Y., E.C.O. and F.G.; writing—review and editing, A.Y., E.C.O. and F.G.; visualisation, A.Y. and E.C.O.; supervision, E.C.O.; project administration, E.C.O.; funding acquisition, E.C.O. All authors have read and agreed to the published version of the manuscript.

Financial support. This study was supported by the Recep Tayyip Erdoğan University, Scientific Research Projects Coordinator Unit (BAP) under Project No. FBA-2025-1947. This article was also supported by the Read and Publish agreement between the Scientific and Technological Research Council of Türkiye (TÜBİTAK) ULAKBİM, under the EKUAL Project, and Cambridge University Press (2026–2028).

Competing interests. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethical standards. The study did not involve human participants, human tissue or vertebrate animal subjects and therefore did not require approval from an institutional review board. The authors acknowledge the use of ChatGPT (GPT-5.5 Thinking, OpenAI) to improve the clarity and fluency of the manuscript. After using this tool, the authors reviewed and edited the content as necessary and take full responsibility for the final content of the publication.

References

[1] Valavanis, K.P. and Vachtsevanos, G.J. *Handbook of Unmanned Aerial Vehicles*, Springer, 2015, Dordrecht.
[2] Aşkan, A., Tangöz, S. and Konar, M. An investigation of aerodynamic behaviours and aerodynamic performance of a model wing formed from different profiles, *Aeronaut. J.*, 2023, **127**, pp 676–697. https://doi.org/10.1017/aer.2022.85
[3] Gencag, M.S., Yildirim Dalkiran, F., Kekec, E.T. and Konar, M. Impact analysis of solar cells on fixed-wing UAV: simulation-based performance evaluation under varying environmental conditions, *Aeronaut. J.*, 2026, **130**, pp 1920–1943. https://doi.org/10.1017/aer.2026.10147
[4] Konar, M., Hatipoğlu, S.A. and Akpınar, M. Improvement of UAV thrust using the BSO algorithm-based ANFIS model, *Aeronaut. J.*, 2024, **128**, pp 2364–2373. https://doi.org/10.1017/aer.2024.30
[5] Konar, M. and Hatipoğlu, S.A. Maximisation the autonomous flight performance of unmanned helicopter using BSO algorithm, *Aeronaut. J.*, 2024, **128**, pp 2656–2667. https://doi.org/10.1017/aer.2024.44
[6] Paul, T.V., Vadakkekara Itty, G. and Indiran, T. Nonlinear dynamic modeling of a quadrotor and its control using sliding mode controller, *J. Aerosp. Eng.*, 2023, **36**, p 04023077. https://doi.org/10.1061/JAEEEZ.ASENG-3722
[7] Ghalamchi, B., Jia, Z. and Mueller, M.W. Real-time vibration-based propeller fault diagnosis for multicopters, *IEEE/ASME Trans. Mechatron.*, 2020, **25**, pp 395–405. https://doi.org/10.1109/TMECH.2019.2947250
[8] EASA. *Easy Access Rules for Unmanned Aircraft Systems (Regulation (EU) 2019/947 and Regulation (EU) 2019/945), Revision September 2022*, European Union Aviation Safety Agency, 2022.
[9] FAA and TSA. Normalizing Unmanned Aircraft Systems Beyond Visual Line of Sight Operations (Notice of Proposed Rulemaking, Part 108), Rulemaking ID 2120-AL82, Federal Register, 7 August 2025.
[10] Saufi, S.R., Ahmad, Z.A.B., Leong, M.S. and Lim, M.H. Challenges and opportunities of deep learning models for machinery fault detection and diagnosis: a review, *IEEE Access*, 2019, **7**, pp 122644–122662. https://doi.org/10.1109/ACCESS.2019.2938227
[11] Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P. and Gao, R.X. Deep learning and its applications to machine health monitoring, *Mech. Syst. Signal Process.*, 2019, **115**, pp 213–237. https://doi.org/10.1016/j.ymssp.2018.05.050
[12] Khan, S. and Yairi, T. A review on the application of deep learning in system health management, *Mech. Syst. Signal Process.*, 2018, **107**, pp 241–265. https://doi.org/10.1016/j.ymssp.2017.11.024

Downloaded from https://www.cambridge.org/core. 18 Nov 2026 at 10:36:50, subject to the Cambridge Core terms of use.

[13] Randall, R.B. and Antoni, J. Rolling element bearing diagnostics: a tutorial, *Mech. Syst. Signal Process.*, 2011, **25**, pp 485–520. https://doi.org/10.1016/j.ymssp.2010.07.017
[14] Liu, R., Yang, B., Zio, E. and Chen, X. Artificial intelligence for fault diagnosis of rotating machinery: a review, *Mech. Syst. Signal Process.*, 2018, **108**, pp 33–47. https://doi.org/10.1016/j.ymssp.2018.02.016
[15] Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N. and Nandi, A.K. Applications of machine learning to machine fault diagnosis: a review and roadmap, *Mech. Syst. Signal Process.*, 2020, **138**, p 106587. https://doi.org/10.1016/j.ymssp.2019.106587
[16] Sharma, V. and Parey, A. A review of gear fault diagnosis using various condition indicators, *Proc. Eng.*, 2016, **144**, pp 253–263. https://doi.org/10.1016/j.proeng.2016.05.131
[17] Rangel-Magdaleno, J., Peregrina-Barreto, H., Ramirez-Cortes, J. and Cruz-Vega, I. Vibration analysis of partially damaged rotor bar in induction motor under different load condition using DWT, *Shock Vib.*, 2016, **2016**, p 3530464. https://doi.org/10.1155/2016/3530464
[18] Iannace, G., Ciaburro, G. and Trematerra, A. Fault diagnosis for UAV blades using artificial neural network, *Robotics*, 2019, **8**, p 59. https://doi.org/10.3390/robotics8030059
[19] Altinors, A., Yol, F. and Yaman, O. A sound based method for fault detection with statistical feature extraction in UAV motors, *Appl. Acoust.*, 2021, **183**, p 108325. https://doi.org/10.1016/j.apacoust.2021.108325
[20] Bondyra, A., Gardecki, S., Gasior, P. and Giernacki, W. Performance of Coaxial Propulsion in Design of Multi-Rotor UAVs, in *Challenges in Automation, Robotics and Measurement Techniques, Advances in Intelligent Systems and Computing*, vol. 440, Springer, 2016, pp 523–531. https://doi.org/10.1007/978-3-319-29357-8_46
[21] Khalastchi, E., Kalech, M., Kaminka, G.A. and Lin, R. Online data-driven anomaly detection in autonomous robots, *Knowl. Inf. Syst.*, 2015, **43**, pp 657–688. https://doi.org/10.1007/s10115-014-0754-y
[22] Breiman, L. Random forests, *Mach. Learn.*, 2001**45**, pp 5–32.
[23] Cortes, C. and Vapnik, V. Support-vector networks, *Mach. Learn.*, 1995, **20**, pp 273–297.
[24] Cover, T. and Hart, P. Nearest neighbor pattern classification, *IEEE Trans. Inf. Theory*, 1967, **13**, pp 21–27.
[25] Quinlan, J.R. Induction of decision trees, *Mach. Learn.*, 1986, **1**, pp 81–106.
[26] Konar, M., Özdemir, D., Fenerci, M. and Erşen, M. Mini UAVs' flight data estimation in navigation phase with LSTM method within the CRISP-DM framework, *Aeronaut. J.*, 2026, **130**, pp 1546–1565. https://doi.org/10.1017/aer.2026.10145
[27] Konar, M., Cam, O. and Aktaş, M.O. BSO algorithm and artificial neural network aided emission optimisation for gas turbine engine, *Aeronaut. J.*, 2025, **129**, pp 107–122. https://doi.org/10.1017/aer.2024.96
[28] Karaburun, N.N., Hatipoğlu, S.A. and Konar, M. Aircraft takeoff speed prediction with machine learning: parameter analysis and model development, *Aeronaut. J.*, 2025, **129**, pp 1534–1549. https://doi.org/10.1017/aer.2024.164
[29] Korkmaz Can, N., Ozkat, E.C., Ceryan, N. and Ceryan, S. Benchmarking ML approaches for earthquake-induced soil liquefaction classification, *Appl. Sci.*, 2025, **15**, p 11512. https://doi.org/10.3390/app152111512
[30] Pourpanah, F., Lim, C.P., Wang, X., Tan, C.J., Seera, M. and Shi, Y. A hybrid model of fuzzy min–max and brain storm optimization for feature selection and data classification, *Neurocomputing*, 2019, **333**, pp 440–451. https://doi.org/10.1016/j.neucom.2019.01.011
[31] Wang, B., Lei, Y., Li, N. and Li, N. A hybrid prognostics approach for estimating remaining useful life of rolling element bearings, *IEEE Trans. Rel.*, 2020, **69**, pp 401–412. https://doi.org/10.1109/TR.2018.2882682
[32] Altunkaya, A.N., Ozkat, E.C. and Avci, M. Analytical-to-AI pipeline: modeling and optimization of entropy generation in pulsating non-Newtonian heat flow, *Comput. Math. Appl.*, 2026, **205**, pp 195–211. https://doi.org/10.1016/j.camwa.2025.12.021
[33] Isbilen, F., Bektas, O. and Konar, M. Deep learning and similarity-based models for predicting turbofan engine remaining useful life: insights from the CMAPSS dataset, *Aeronaut. J.*, 2025, **129**, pp 2004–2035. https://doi.org/10.1017/aer.2025.25
[34] Isbilen, F., Bektas, O., Avsar, R. and Konar, M. Improved machine learning models with a similarity-based approach for remaining useful life prediction, *Aeronaut. J.*, 2025, **129**, pp 485–505. https://doi.org/10.1017/aer.2024.101

Cite this article: Yilmaz A., Ozkat E.C. and Gul F. A low-cost wireless accelerometer approach to UAV propeller fault detection. *The Aeronautical Journal*, https://doi.org/10.1017/aer.2026.10235

Downloaded from https://www.cambridge.org/core. 18 Nov 2026 at 10:36:50, subject to the Cambridge Core terms of use.