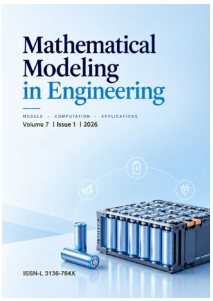
UNIVERSITY OF WEST ATTICA

eISSN: 3136-764X

Mathematical Modeling in Engineering

Vol. 7, No. 1, January, 2026



MambaVision Model for UAV Propeller Damage Identification Driven by Acoustic and Vibration Flight Data

Mateusz Kozak[a], Mariusz Jacek Kijewski[b], Julia Karwowska[a,*]

[a] *Faculty of Information Technology and Computer Science, Lodz University of Technology, Lodz, 90-924, Poland*

[b] *Faculty of Information Technology, Polish-Japanese Academy of Information Technology, Warsaw, 02-007, Poland,*

Received 18 November 2025; Revised 05 February 2026; Accepted 10 February 2026; Published 16 February 2026

Abstract

Small unmanned aerial vehicles increasingly operate near people, infrastructure, and dense sensor networks, yet propeller damage remains difficult to identify during flight because acoustic radiation and structural vibration vary with rotor speed, payload, maneuver mode, and air disturbance. This paper proposes an acoustic and vibration flight-data-driven MambaVision model for propeller damage identification. Synchronized microphone and accelerometer measurements are transformed into acoustic spectrogram tokens and vibration order-map tokens, while rotor speed and operating descriptors are retained to reduce flight-condition interference. The model combines local visual attention, selective state-space scanning, and gated cross-modal fusion so that short-range time-frequency damage patterns and longer flight-window dependencies can be learned jointly. Five propeller states are considered, including healthy blades, leading-edge nicking, surface erosion, root cracking, and imbalance deformation. The proposed method achieved 97.1% macro F1, 96.8% macro recall, and 97.0% overall accuracy on held-out flight sessions. Under added payload and crosswind conditions, the recognition accuracy remained above 94%, and ablation results showed that multimodal fusion and state scanning improved both accuracy and convergence stability. These results indicate that acoustic-vibration fusion with MambaVision can support practical UAV propeller health monitoring, especially when in-flight damage signatures are weak, coupled, and affected by changing operating conditions. The framework is suitable for onboard maintenance screening.

Keywords: Propeller Damage Identification, Acoustic Vibration Fusion, Flight Data, MambaVision, Unmanned Aerial Vehicles

1. Introduction

Multirotor unmanned aerial vehicles have left the controlled test area to conduct inspection and mapping work, as well as emergency logistics, and propulsion health is now a safety issue for daily operations [1]. A light-weight rotating component, a small geometric imperfection in the propeller can cause thrust asymmetry, periodic vibration of the airframe, and a change in the broadband acoustic field around the rotor disk [2]. Based on field investigations, the causes of propeller damage are often dust and rain erosion, minor ground contact, and repeated transportation impacts [3]. If the blade surface is asymmetric, then a closed-loop speed control system will amplify the fluctuations in induced load when the controller increases the torque of the motor to compensate for thrust loss [4]. Therefore, acoustic monitoring has been used to detect sidebands and tonal shifts caused by damaged blades during flight without interrupting the flight [5]. Vibration sensing is also used to monitor the vibration of hub-order components and fuselage resonance that cannot be detected through airborne sound [6]. Both channels are affected by rotor speed, payload, flight mode and wind-induced inflow; therefore, a diagnostic system trained only on bench data may have poor performance in a real-aircraft

* Corresponding author: julia.ka@p.lodz.pl

https://doi.org/10.68406/mme.2026.vol7iss1nm6:68-79

➢ This is an open access article under the CC BY-NC-SA 4.0 license (https://creativecommons.org/licenses/by/4.0/).

© Authors retain all copyrights.

environment. According to the newest analysis of flight data, only after integrating sensing, synchronization and model design will the damage signatures be accurate.

At present, the two general types of propeller diagnosis methods are handcrafted feature pipelines and deep representation learning. Al-Haddad et al. [7] used non-traditional complexity and energy features to support UAV propeller fault diagnosis, showing that signal descriptors can provide useful diagnostic information before deep representation learning is introduced. The aforementioned method is relatively easy to understand, but the feature threshold usually optimizes a small operating window and performs poorly when the rotor speed changes rapidly during maneuvering [8]. Convolutional neural networks reduce the requirement for manual feature engineering by learning spectrogram textures, but their local receptive fields cannot capture long-range temporal changes in the state of imbalance and intermittent blade loading [9]. Transformers improve the modelling of global context, but their quadratic attention cost is not suitable for high-rate acoustic vibration streams on embedded UAV hardware [10]. Recently, state-space sequence models and newer Mamba-type architectures have also proposed another way to address the problem of long sequences by introducing input-dependent dynamics and linear complexity, and are sensitive to local variations [11]. Based on the above attributes, a multi-modal model has been constructed to process acoustic spectrograms and vibration order maps as auxiliary images with strong temporal correlations.

The primary problem this paper aims to solve is how to label propeller damage in flight data that also contains information such as the source of the signal, operating conditions and changes in mechanical response. A propeller nick is generally a weak high-frequency acoustic sideband, and root cracking is shown as an increase in vibration at the blade-passing frequency and its harmonics. Surface erosion spreads the energy in the mid-frequency band and does not have an independent peak. If both the amplitude of vibration and asymmetry in acoustics are neglected, unbalanced deformation will be mistaken for a normal load shift. Based on the above observations, a good model should have local visual patterns, long-sequence dependencies, and sensor-specific evidence, and should avoid forcing all modalities into a single flattened descriptor. MambaVision is suitable for this purpose, and it can maintain the structure of visual patches and model ordered flight windows through selective state propagation.

Against this background, this paper develops a MambaVision model that learns from synchronized acoustic and vibration flight data, links the learned evidence to physically meaningful propeller damage mechanisms, and evaluates the method on five controlled propeller states under varying maneuver and operating conditions. The first section introduces the engineering background and the diagnostic gap. Section 2 reviews acoustic monitoring, vibration analysis, multimodal learning, and state space recognition methods. Section 3 presents the acquisition pipeline, token construction, MambaVision architecture, and training objective. Section 4 reports the experimental dataset, comparative performance, robustness analysis, ablation study, interpretability evidence, and deployment trade off. Section 5 concludes the work by describing the main contribution, current limitations, and future research directions.

2. Related Work

2.1. Acoustic and Vibration Based Propeller Diagnosis

Acoustic inspection of rotating machinery is used to detect damage by analyzing changes in pressure fluctuations and turbulent wakes caused by motion. Acoustic fields around UAV propellers are of great interest, and blade-passage tones, sideband modulation and other components of broadband noise all depend on blade geometry and rotor speed [12]. The first studies used microphones in anechoic or semi-anechoic spaces, where clean spectra could be obtained to identify unbalanced or chipped propellers by means of tonal energy ratios. In actual flight, the above assumption is no longer valid because of motor noise, aerodynamic interference and other sounds in the environment. Therefore, researchers have added short-time Fourier transforms, Mel-frequency descriptors and adaptive spectral normalization to stabilize the acoustic features of different flight modes [13]. Although the above methods can find some useful patterns, they generally treat consecutive windows as independent and therefore lose the ability to capture transient phenomena in acceleration and yaw.

Vibration-based propeller diagnosis is grounded in rotor dynamics and structural transmission. A damaged blade changes mass distribution and aerodynamic loading, causing order components to appear in accelerometer channels near the motor arm or fuselage [14]. Compared with acoustic sensing, vibration sensing is less sensitive to ambient sound and can identify low-frequency imbalance more directly. However, fuselage resonance, controller action, and landing-gear flexibility can weaken blade-specific evidence. Therefore, multisensor strategies are needed to preserve the complementary physical meanings of acoustic radiation and mechanical excitation [15].

2.2. Deep Sequence Models and Multimodal Recognition

Deep learning uses learned representations to address the problem of fault diagnosis without hand-crafted descriptors. Convolutional Networks are relatively good at detecting local changes in texture of spectrograms and vibration maps caused by damage, and they are still widely used due to their training stability [16]. Flight signals also have issues with

long-range dependence; for example, the history of speed acceleration, periodic load transfer, and the delayed vibration growth after maneuvers are all manifestations of this dependence. Recurrent networks and temporal convolutional networks can address the problem of dependency but may lose some fine local details or need to design dilation carefully. Transformers consider the whole world using attention and have achieved excellent recognition results for image-like signal representations [17]. However, the cost of attention and data requirements for UAV datasets is relatively high; generally, they are smaller than those of general computer vision datasets and limited by safety tests.

Another kind of sequence model is a selection of state-space models with input-dependent parameters for the hidden state. Mamba-based models have attracted the attention of researchers due to their long-context capabilities and linear-time scanning and efficient memory use [18]. MambaVision extends the previous idea by introducing local visual processing and state-space scanning to address the problem of time-frequency signal images in visual recognition. The first problem of multi-modal damage recognition is that it often only concatenates acoustic and vibration patches; otherwise, both modalities can keep their original physical meanings and share evidence at the location of the overlapping damage signature. Cross-modal gating, modality-aware normalization and late feature pooling have been proposed in the field of related monitoring [19]. In the same direction, this paper will also study damage to UAV propellers during flight tests and change operating conditions for engineering reliability evaluation.

3. Acoustic Vibration MambaVision Framework

3.1. Flight Data Acquisition and Multimodal Token Construction

The first module of the framework synchronously collects airborne acoustic pressure, triaxial vibration, rotor speed, throttle command, and flight attitude. The microphone captures aerodynamic radiation and blade-passage sidebands, while the accelerometer records hub-transmitted order components and structural vibration. Rotor speed and flight attitude are retained as operating descriptors so that the subsequent normalization stage can reduce speed-dependent frequency drift without removing damage-sensitive modulation.

Figure 1 shows the acquisition and preprocessing flow used to transform raw flight signals into synchronized acoustic spectrogram tokens, vibration order tokens, and operating condition descriptors before model inference.

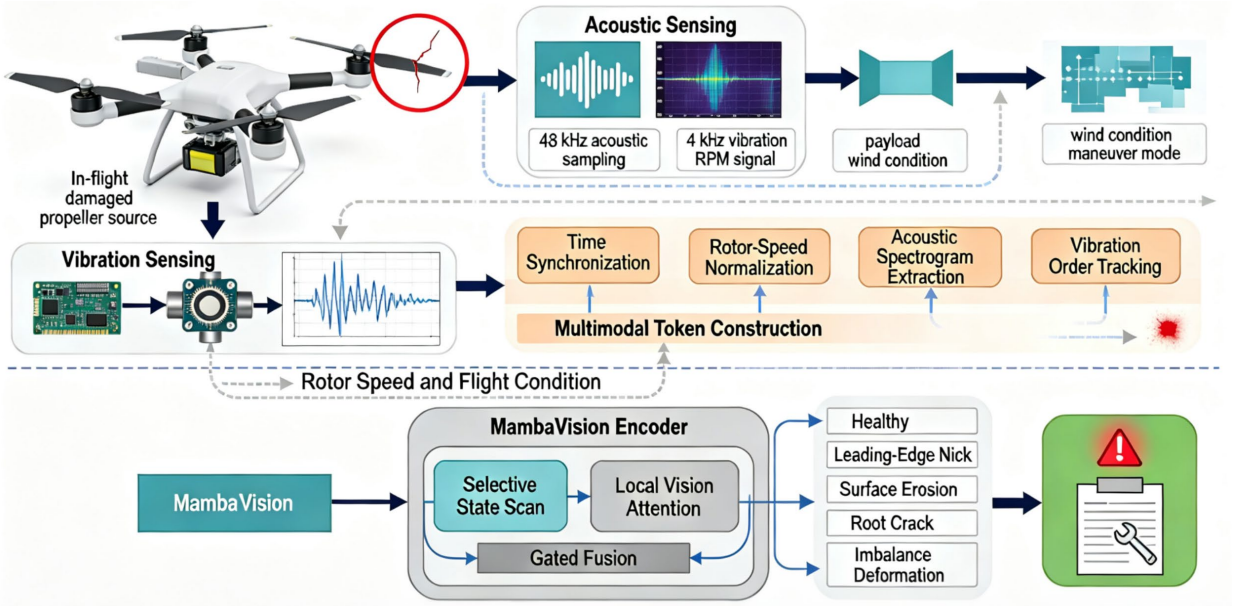


Figure 1. Acoustic and vibration flight data pipeline for UAV propeller damage identification

The two synchronized signal images used by the model are defined as follows.

$$x_a(t) = log(1 + |STFT(p(t))|) \tag{1}$$

The vibration order map is then obtained from the synchronized acceleration channels and rotor speed.

$$x_v(t) = OrderMap\big(a_x(t), a_y(t), a_z(t), r(t)\big) \tag{2}$$

Here, $p(t)$ denotes acoustic pressure, $a_x(t), a_y(t)$, and $a_z(t)$ denote vibration acceleration along the three axes, and $r(t)$ denotes rotor speed. The acoustic and vibration maps are resized into equal patch grids, then projected into tokens with modality specific linear layers so that local physical structures are not lost before fusion. The initial multimodal token sequence is constructed as follows.

$$z_0 = [E_a(x_a) + e_a, E_v(x_v) + e_v, E_o(o)] \tag{3}$$

3.2. MambaVision Encoder and Cross Modal Fusion

The encoder has parallel shallow embedding layers and a shared stack of local visual attention blocks and selective state space blocks. Local attention extracts short-range damage features, such as narrow sidebands, harmonic broadening and order-ridge thickening. The state space branch scans the ordered token sequence to model the temporal relationship of adjacent flight windows. A Gated Fusion Unit gives different weights to the evidence of acoustics and vibration in different operating conditions. Acoustic features are more likely to occur in the case of blade-surface erosion, and vibrations caused by root cracking or imbalance deformation are relatively prominent.

Figure 2 is the structure of the selected model, and a gated fusion layer integrates local visual attention and selective state scanning prior to damage-aware feature pooling.

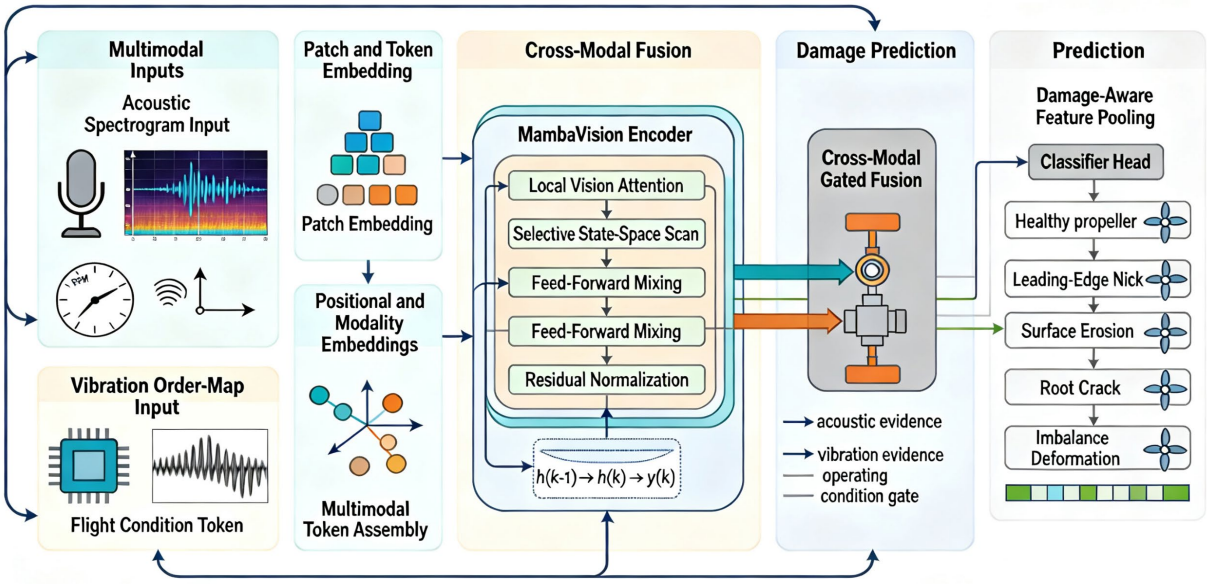


Figure 2. MambaVision architecture with selective state scanning and gated acoustic vibration fusion

The selective state update used to scan ordered flight tokens is expressed as follows.

$$h_k = A_k(o)h_{k-1} + B_k(o)z_k, y_k = C_k(o)h_k + Dz_k \tag{4}$$

The matrices A_k, B_k, and C_k are conditioned by operating descriptors o, allowing the state update to adapt to rotor speed, payload, and maneuver condition. The acoustic and vibration streams are fused with a gate that is learned from both feature channels rather than fixed by manual weighting. The fusion operation is written as follows.

$$g = \sigma\big(W_g[y_a; y_v; o]\big), f = g \odot y_a + (1 - g) \odot y_v \tag{5}$$

3.3. Training Objective and Inference Rule

The classifier receives the fused feature vector and outputs the probability of five propeller states: healthy, leading-edge nick, surface erosion, root crack, and imbalance deformation. Training uses weighted cross entropy to reduce bias toward frequently sampled healthy windows and a compactness regularizer that encourages stable embeddings for repeated flights of the same damage type. The model is optimized with balanced mini batches across damage classes and flight regimes. During inference, window level predictions are aggregated across a short flight segment, while the maximum damage probability is retained for maintenance alarm generation. The training objective is defined as follows.

$$L = -\sum_{c=1}^{C} w_c q_c log(p_c) + \lambda \left\| f_i - \mu_{q_i} \right\|_2^2 \tag{6}$$

To avoid using a single noisy window to determine the final result, the inference rule takes the mean probability and persistent alarm evidence of neighboring windows. This rule is more suitable for propeller damage; otherwise, genuine faults will occur repeatedly during rotor rotation, and isolated acoustic disturbances will not be uniform in adjacent sections.

4. Experimental Validation and Flight Data Analysis

4.1. Dataset and Experimental Setting

A flight dataset was collected using a 450 mm-class quadrotor under five propeller states rather than five propeller speeds. The classes were healthy, leading-edge nick, surface erosion, root crack, and imbalance deformation. The leading-edge nick contained a 3 mm local cut near the outer span, surface erosion used distributed roughening along the outer chord, the root crack class used a controlled 5 mm root crack, and the imbalance class used a trailing-edge mass offset with chordwise bending. A similar controlled-damage protocol has been used in the field of rotor-monitoring research to differentiate between measurable degradation and sudden failure [20].

Table 1 reports five propeller states with at least 1012 windows per class, 17 to 18 flight sessions per state, and mean rotor speeds between 4690 and 4815 rpm, providing enough operating diversity to evaluate damage identification beyond stationary bench measurements.

Table 1. Flight dataset composition and dominant damage signatures

Propeller state	Damage description	Flights	Windows	Mean RPM	Primary signature
Healthy	Balanced undamaged propeller	18	1040	4720	Low sideband energy
Leading edge nick	3 mm cut near outer span	18	1032	4765	High frequency acoustic sideband
Surface erosion	Distributed roughened chord	18	1018	4690	Broadband spectral thickening
Root crack	5 mm stabilized root crack	17	1012	4815	Strong 1P vibration rise
Imbalance deformation	Mass offset and chord bending	18	1082	4780	Order ridge asymmetry

Raw acoustic signals were sampled at 48 kHz, and vibration signals were sampled at 4 kHz. Both streams were aligned using controller timestamps. Window labels were assigned at the flight-session level and verified through pre-flight blade inspection. Acoustic amplitudes were normalized by the median background level, and vibration magnitudes were normalized according to motor-arm calibration. Flight sessions, rather than adjacent windows, were used for validation splitting to reduce leakage between training and testing data [21]. All types of damage and at least three motion modes were included in each mini-batch for training. Add a small amount of low-frequency noise, rotor-speed jitter and minor vibration scaling to the data to simulate measurement errors without causing artificial damage [22].

Figure 3(a) shows that rotor speed mainly reduced the margin of the more severe damage classes rather than causing a uniform collapse across all states. Healthy windows decreased from 98.5% at 3200 rpm to 96.1% at 6200 rpm, while root crack decreased from 96.2% to 92.7%, indicating that high speed loading made crack evidence less stable in short windows. Figure 3(b) separates the same flight conditions in a coupled acoustic vibration space: root crack reached a 7.3 dB sideband ratio and 1.18 g vibration RMS, whereas imbalance reached 6.9 dB and 1.04 g. The two damaged states both exceeded the 0.9 g vibration threshold, but their acoustic positions remained different, which explains why a single vibration threshold would not be sufficient for class level identification.

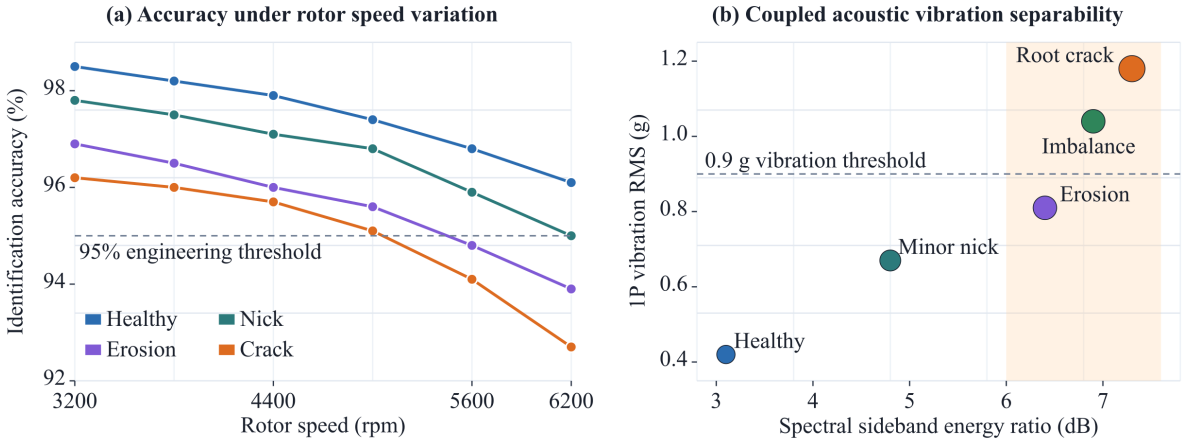


Figure 3. Flight signal quality and separability: (a) accuracy under rotor speed variation; (b) coupled acoustic vibration separability

4.2. Damage Classification Performance

First, the classification performance was tested on the held-out flight sessions. The MambaVision model achieved 97.1% macro F1, 96.8% macro recall, and 97.0% overall accuracy. Locating the root crack was difficult; sometimes, the first crack window appeared as a surface erosion in the acoustic spectrum, but vibration-order evidence helped to rule out this possibility. The main kinds of recognition errors are erosion and nicking, and these errors are related to high-frequency components in the sound [23]. As shown in the confusion matrix, healthy windows were rarely identified as damaged; only three false alarms occurred among the 216 held-out healthy windows. A relatively low false-alarm rate is required for the model's recommendations to be implemented by maintenance personnel [24].

Figure 4(a) shows that the confusion matrix was strongly diagonal, with 213 correctly identified healthy windows, 201 leading edge nick windows, 194 erosion windows, 188 root crack windows, and 197 imbalance windows. The off diagonal entries were not random: erosion was confused with nicking in nine windows, and crack was confused with erosion in four windows, matching the overlap between high frequency acoustic roughness and early structural damage. Figure 4(b) compares recognition metrics across five models and shows that MambaVision achieved 97.1% macro F1 and 96.8% macro recall. The gain over ViT was 1.9 percentage points in macro F1, while the gain over CNN was 4.7 percentage points, indicating that long range state scanning contributed beyond local spectrogram texture learning.

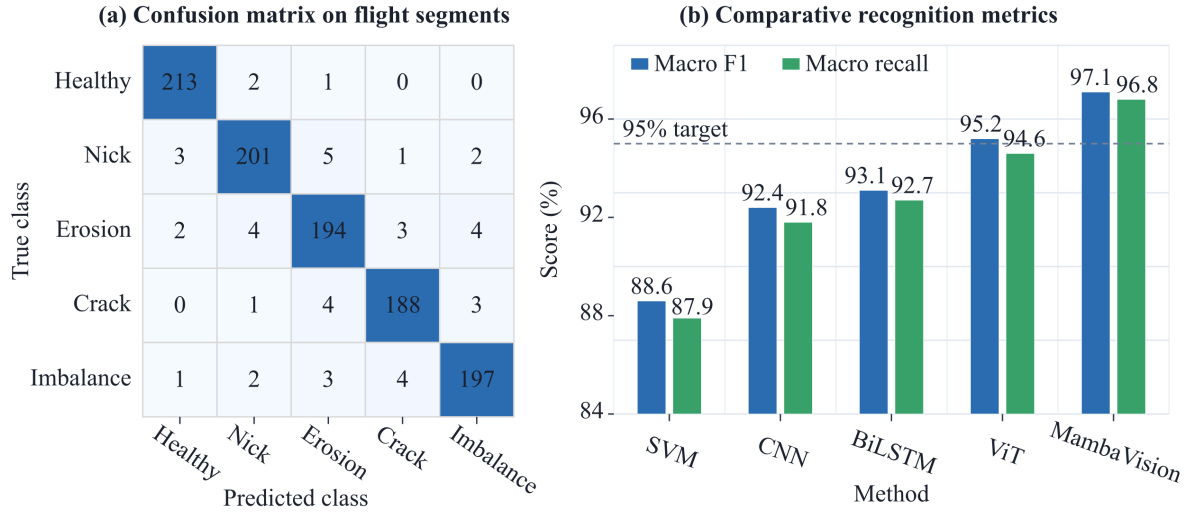


Figure 4. Classification performance: (a) confusion matrix on flight segments; (b) comparative recognition metrics

The comparison demonstrates that model structure matters as much as input modality. Classical spectral and order descriptors are widely used because they remain physically interpretable under controlled rotor conditions [25]. The SVM baseline used these descriptors and achieved 88.6% macro F1 in the present test. The CNN baseline improved to 92.4% macro F1 by learning local time frequency textures, but it still confused erosion and nicking when rotor speed shifted. BiLSTM achieved 93.1% macro F1 because it used temporal context, although it lacked strong spatial modeling for spectrogram and order maps. ViT achieved 95.2% macro F1 but required heavier computation and showed weaker recall for short root crack events. MambaVision benefited from the selective state scan and local visual blocks, giving the best balance between recognition accuracy and computational demand.

Figure 5(a) reports the payload sensitivity of healthy, root crack, and imbalance classes. Root crack accuracy declined to 94.5% at 450 g added payload, which is larger than the healthy class decline because payload induced speed compensation partly masked the crack related order component. Figure 5(b) shows the acoustic noise response through distributional rather than single point values; the median segment accuracy stayed near 96% at 15 dB SNR but dropped to 91.9% at 5 dB SNR, and the wider violin shape indicates that some windows became unstable when broadband noise overlapped damage sidebands. Figure 5(c) gives a crosswind trend from 97.4% at 0 m/s to 94.6% at 8 m/s, with visibly larger uncertainty at the strongest wind speed. Figure 5(d) compares maneuver regimes and shows that MambaVision retained 95.4% accuracy during yaw, whereas the CNN baseline fell to 87.6%, reflecting the benefit of modeling temporal state continuity during attitude changes.

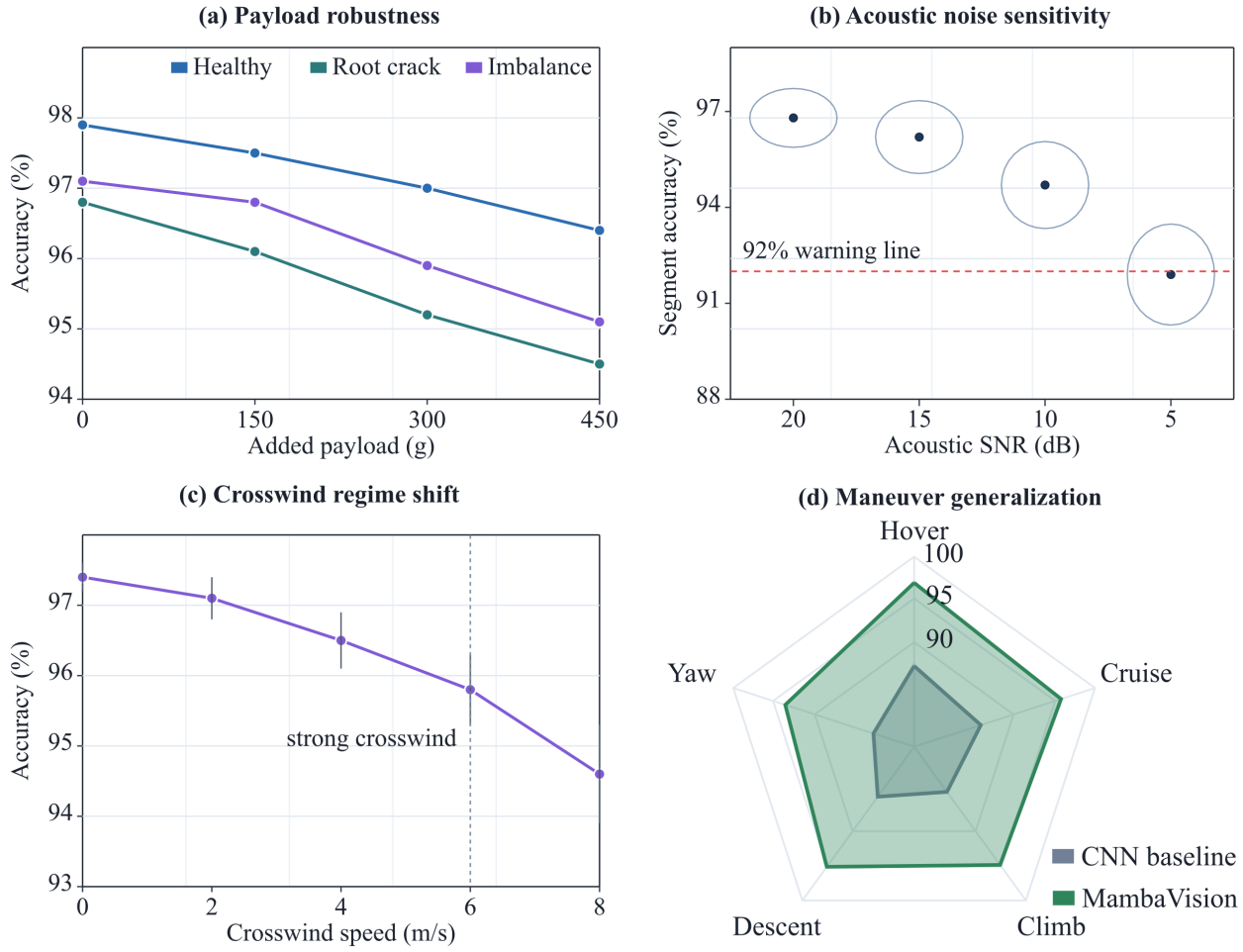


Figure 5. Robustness analysis: (a) payload robustness; (b) acoustic noise sensitivity; (c) crosswind regime shift; (d) maneuver generalization

Robustness tests were conducted to isolate the effect of damage from that of operation. Payload addition reduced accuracy because controller compensation changed the distribution of rotor speed and obscured small acoustic sidebands. The main sources of acoustic noise are nicking and erosion; they are generally high-frequency noises. Crosswind had a larger range for both sound pressure and vibration; however, after rotor speed normalization and state scanning, the model could still distinguish between continuous damage and fluctuating inflow reasonably well. Comparable multimodal diagnostic studies have also found that only a small number of training data are used and are thus less sensitive to unobserved operating conditions [26].

4.3. Ablation, Interpretability, and Deployment Analysis

Ablation experiments determined the effects of the modules. The accuracy of the acoustic-only input was 90.7%, and that of the vibration-only input was 91.9%; therefore, neither channel could identify all kinds of damage on its own. Simple feature fusion achieved a relatively high accuracy of 94.8%, but without the selective state mixer, there was a lack of temporal stability in rotor acceleration. For online UAV monitoring, the inference time should be much shorter than the acquisition interval to avoid the accumulation of delayed alarms during flight [27]. A total of 97.1% accuracy and an average inference speed of 12.6ms per window were achieved by the model on embedded GPUs.

Figure 6(a) shows the cost of each architectural choice. Acoustic only input reached 90.7% accuracy and vibration only input reached 91.9%, confirming that each sensor missed part of the damage mechanism. Feature fusion increased accuracy to 94.8%, and the full MambaVision configuration reached 97.1% with 12.6 ms latency, which means the additional fusion and state scan cost remained small relative to the one second analysis window. Figure 6(b) shows that the full model converged toward approximately 0.10 validation loss after about 35 epochs, while the model without gated fusion stayed higher and the model without state mixing decreased more slowly. The convergence behavior indicates that the proposed modules improved optimization stability as well as final classification accuracy.

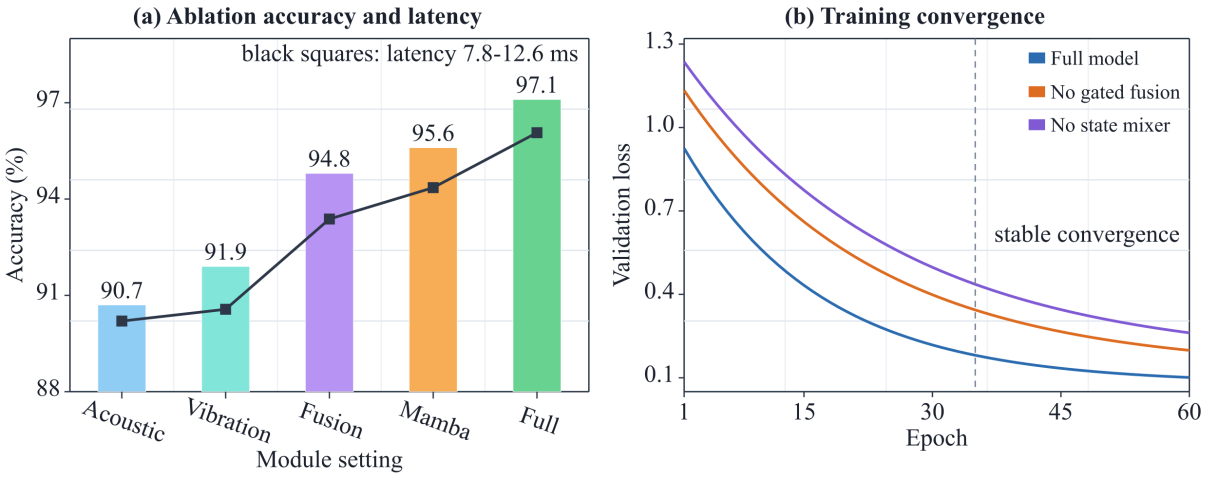


Figure 6. Ablation and convergence: (a) ablation accuracy and latency; (b) training convergence

Latent feature projection, modality attention, segment-level probability and class-specific saliency were used for interpretation analysis. Latent Projection successfully separated the healthy, crack and imbalance samples, and nicking and erosion appeared near the location of changes in high-frequency acoustic radiation. With the expansion of the range of damage caused by combined acoustic vibration, a weak acoustic signal is no longer sufficient. The segment-level crack probability of the eight adjacent windows of a representative flight increased from 0.12 to 0.89, and it did not reach the alarm threshold until many times. Persistence is necessary to address recurrent fault signatures in flight monitoring and rule out random sound artifacts [28].

Figure 7(a) shows that the learned feature space separated healthy, crack, and imbalance states clearly, while nicking and erosion remained closer because both defects produced surface related acoustic changes. Figure 7(b) shows that vibration attention rose to 0.61 at severity grade 4, whereas acoustic attention reached 0.55, suggesting that higher severity shifted the model toward mechanical evidence without discarding acoustic information. Figure 7(c) presents a representative crack sequence in which the probability crossed the 0.65 alarm threshold at segment 5 after increasing across earlier windows, so the alarm depended on persistent evidence rather than one isolated signal burst. Figure 7(d) shows class specific saliency gains, with root crack reaching 6.7 for acoustic evidence and 4.9 for vibration evidence, while erosion and nicking placed greater relative emphasis on acoustic saliency.

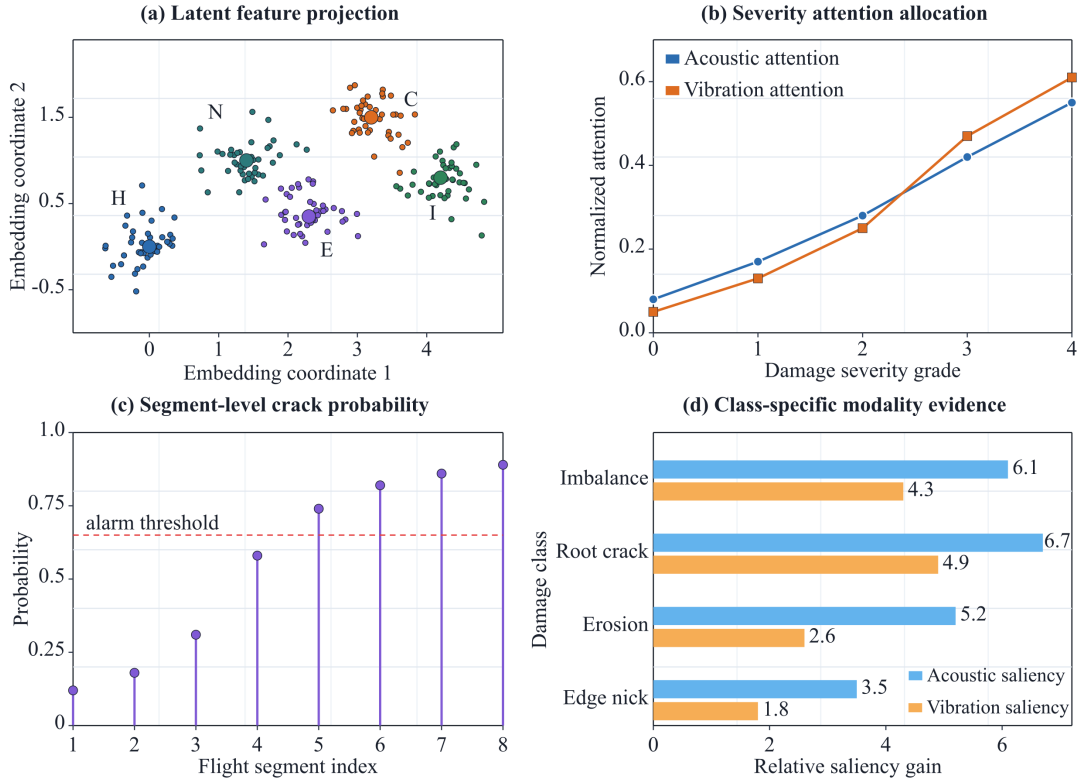


Figure 7. Interpretability evidence: (a) latent feature projection; (b) severity attention allocation; (c) segment level crack probability; (d) class specific modality evidence

Deployment analysis considered recognition accuracy, latency, memory use, and processed windows per second. A 1.0 s window offered the best operating point, with 97.1% accuracy and 12.6 ms latency. Shorter 0.25 s windows reduced latency to 5.4 ms but lost periodic context, decreasing accuracy to 93.1%. Longer 2.0 s windows did not improve accuracy and increased latency to 23.4 ms. Hardware tests indicated 168 processed windows per second on Jetson Orin and 52 windows per second on Jetson Nano, both above the real time floor of 30 windows per second. Model pruning and distillation are commonly required when deep diagnostic networks are transferred from embedded GPU platforms to smaller microcontrollers [29]. On the tested flight microcontroller, throughput fell to 19 windows per second, indicating that direct MCU deployment would need compression. Edge deployment remains attractive because onboard detection avoids transmitting high rate acoustic and vibration streams during routine inspection flights [30].

Figure 8(a) identifies the 1.0 s window as the selected operating zone because it achieved 97.1% accuracy with 12.6 ms latency. The 0.25 s window reduced latency to 5.4 ms but accuracy fell to 93.1%, showing that too little periodic context weakened damage recognition; the 2.0 s window increased latency to 23.4 ms without improving accuracy. Figure 8(b) places the method in a deployment cost envelope. Jetson Nano processed 52 windows per second at 396 MB peak memory, which stayed above the 30 windows per second real time floor, while the flight microcontroller reached only 19 windows per second at 182 MB. This gap indicates that embedded GPU deployment is practical for the current model, whereas direct microcontroller deployment would require compression of the fusion and state scan blocks.

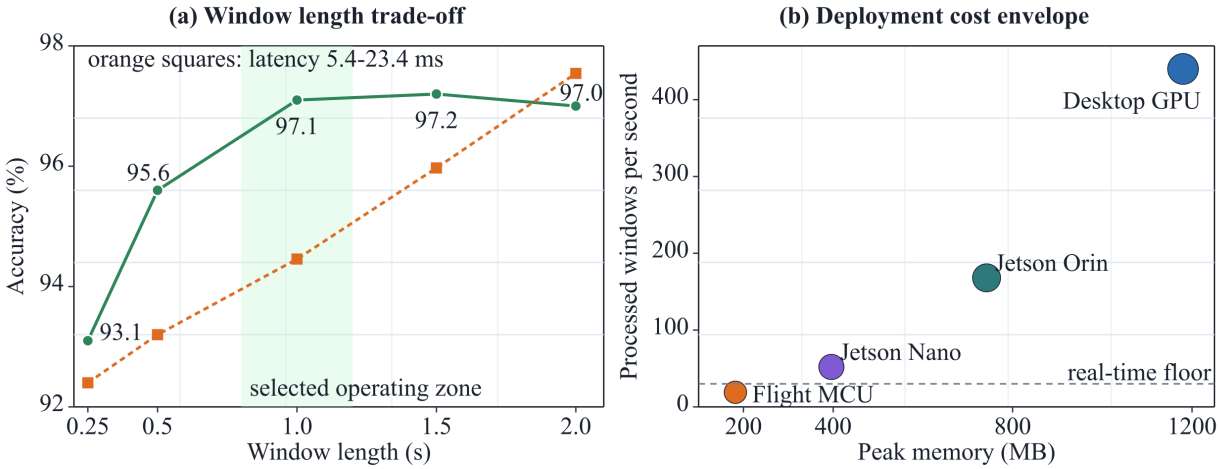


Figure 8. Deployment trade off: (a) window length trade off; (b) deployment cost envelope

Most remaining errors were concentrated between physically similar damage states rather than being randomly distributed across labels. Nicking and erosion both alter blade-surface roughness, so their high-frequency acoustic sidebands may overlap at elevated rotor speeds. Root-crack errors appeared mainly in lower-speed windows where the one-per-revolution vibration component was weaker. These patterns indicate that the classifier responds to interpretable signal mechanisms instead of isolated numerical artifacts.

The deployment results indicate a practical boundary for the proposed method. Embedded GPU hardware can process the selected one-second window with sufficient margin for online monitoring, whereas the tested flight microcontroller lacks adequate throughput before model compression. A realistic deployment path is therefore to prune or distill the gated fusion and selective state-scan blocks while preserving both acoustic and vibration sensing channels.

5. Conclusion

Acoustic and vibration flight data-driven MambaVision model for UAV propeller damage detection was proposed in this paper. The way to obtain synchronized microphone and accelerometer data, normalize rotor-speed-scaled token construction, selective state-space scanning, local visual attention, and gated cross-modal fusion. The built-in framework links the physical damage modes of propellers with an applied recognition architecture for flight data, and experimental analysis has shown that acoustic and vibration evidence are complementary in all types of damage.

Several limitations remain. The flight experiments used controlled damage geometries and a single UAV size class, so the method still requires validation across different propeller materials, blade diameters, motor types, and airframe configurations. Severe weather, dense urban acoustic interference, and naturally aged propellers were not fully represented. Future work should therefore emphasize cross-platform testing and uncontrolled outdoor operating conditions.

More UAVs, older-generation propellers and complex outdoor environments will be added to the dataset in the future. Compression of the model, self-supervised pre-training, uncertainty estimation and continual adaptation will be used to

enhance deployment efficiency and field reliability. A maintenance-oriented alarm strategy will be used to link the predicted damage status with the remaining safe-of-operation time of the actual fleet.

6. Declarations

6.1 Author Contributions

Mateusz Kozak contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. Mariusz Jacek Kijewski contributes to validation, analysis, investigation, data collection, draft preparation. Julia Karwowska contributes to conceptualization, validation, analysis. All authors have read and agreed with the manuscript before its submission and publication.

6.2 Funding and Acknowledgments

This research received no specific financial support from any funding agency.

6.3 Institutional Review Board Statement

Not applicable.

6.4 Informed Consent Statement

Not applicable.

6.5 Data Availability

Not applicable.

6.6 Declaration of Competing Interest

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.

6.7 Declaration of generative AI and AI-assisted technologies

During the preparation of this manuscript, the authors utilized Grammarly for linguistic polishing, logical organization and draft revision. The generative AI tool was not involved in research design, data collection, data analysis, or the equation of core conclusions. All outputs generated by AI were reviewed, revised and verified manually by the authors. The authors take full responsibility for all content of this manuscript.

7. References

[1] Steinhoff, L., Koschlik, A. K., Arts, E., Soria-Gomez, M., Raddatz, F., & Kunz, V. D. (2024). Development of an acoustic fault diagnosis system for UAV propeller blades. CEAS Aeronautical Journal, 15(4), 881-893. https://doi.org/10.1007/s13272-024-00752-8

[2] Rao, M. E., Simon, J., Moll, J., & Schütz, M. F. (2026). Real-time onboard propeller fault diagnosis of autonomous delivery drones through vibration analysis. Structural Health Monitoring, 25(4), 2739-2750. https://doi.org/10.1177/14759217251327224

[3] Kurdi, S. T., Al-Haddad, L. A., Ismail, M. A., Kraiem, H., & Flah, A. (2026). An explainable stacking ensemble for UAV propeller fault diagnosis using electromechanical flight telemetry. Array, 31, 101183. https://doi.org/10.1016/j.array.2026.101183

[4] He, Y., Fang, H., Yan, J., Yang, C., & Zhai, Y. (2025). Computationally efficient UAV fault diagnosis with adaptive vibration denoising: A signal processing approach for rotorcraft systems. Mechanical Systems and Signal Processing, 240, 113413. https://doi.org/10.1016/j.ymssp.2025.113413

[5] Cinoğlu, B. (2024). Acoustic-based diagnostics for UAV propeller damage using HNR and Gaussian Naive Bayes. Aircraft Engineering and Aerospace Technology, 96(7), 972-982. https://doi.org/10.1108/aeat-05-2024-0155

[6] Cinoğlu, B., & Durak, U. (2024). Thrust-level dependent vibration diagnostics of UAV propeller using fast Fourier transform and K-nearest neighbour. International Journal of Sustainable Aviation, 10(4), 297-314. https://doi.org/10.1504/ijsa.2024.142568

[7] Al-Haddad, L. A., Giernacki, W., Basem, A., Khan, Z. H., Jaber, A. A., & Al-Haddad, S. A. (2024). UAV propeller fault diagnosis using deep learning of non-traditional χ2-selected Taguchi method-tested Lempel–Ziv complexity and Teager–Kaiser energy features. Scientific Reports, 14(1), 18599. https://doi.org/10.1038/s41598-024-69462-9

[8] Wen, W., Zhong, W., Liu, Y., Li, X., & Lan, H. (2026). Fault Diagnosis of UAV Rotor Systems Based on Drone Nest Vibration

Analysis. Drones, 10(6), 424. https://doi.org/10.3390/drones10060424

[9] Lalem, M. S. E. I., Ouadah, M., & Touhami, O. (2025). Real-Time Drone Propeller Fault Detection Using Onboard Vibration Sensors and Optimized Machine Learning. Journal of Vibration Engineering & Technologies, 13(8), 547. https://doi.org/10.1007/s42417-025-02122-y

[10] Mohd Ghazali, M. H., Tamilselvam, D., & Rahiman, W. (2025). Comparative analysis between data-driven and visual-based approach in drone propeller imbalance detection. Journal of Vibration and Control, 31(5-6), 822-835. https://doi.org/10.1177/10775463241234158

[11] Sehri, M., Yan, T., Chauhan, S., & Vashishtha, G. (2026). Temporal Domain Vibration Fault Diagnosis of Drone Blades via Selective Embedding. Machines, 14(2), 241. https://doi.org/10.3390/machines14020241

[12] Iqbal, M., Madan, A. K., & Ahmad, N. (2024). Vibration and acoustic signal-based bearing fault diagnosis in CNC machine using an improved deep learning. Iran Journal of Computer Science, 7(4), 723-733. https://doi.org/10.1007/s42044-024-00205-9

[13] Adaika, Z., & Boumehraz, M. (2026). Lightweight deep neural network based fault detection and diagnosis for quadrotor propeller faults. International Journal of Intelligent Robotics and Applications, 10(1), 259-272. https://doi.org/10.1007/s41315-025-00500-z

[14] Makhtar, S. N., Mohd Sabudin, U. S., Zaini, M. H., Bohari, B., Ahmad, K. A., & Kurnianingsih, K. (2025). Propeller Fault Classification for Quadrotor Aerial Vehicle using Spatial Displacement Statistical Features. Journal of Advanced Research in Applied Sciences and Engineering Technology, 65(2), 210-221. https://doi.org/10.37934/araset.65.2.210221

[15] Salahudden, S., & Dwivedi, V. S. (2026). Neural network based fault identification for quadrotor thrust loss and propeller misalignment. Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering, 240(6), 873-886. https://doi.org/10.1177/09544100251405001

[16] Wu, Y., Xie, T., & Yang, S. (2026). Fault diagnosis method for quadrotor UAV based on ResCNN-LSTM-attention. Engineering Research Express, 8(16), 165331. https://doi.org/10.1088/2631-8695/ae9a37

[17] Rahmati, M., & Rahmati, N. (2025). A multimodal deep learning framework for real-time defect recognition in industrial components using visual, acoustic and vibration signals. Journal of Intelligent Manufacturing and Special Equipment, 6(3), 273-292. https://doi.org/10.1108/jimse-07-2025-0015

[18] Zhang, W., Tong, J., Liao, F., & Zhang, Y. (2026). Simulation-to-Reality Quadrotor Propeller Damage Detection in Windy Environments. Unmanned Systems, 14(04), 749-760. https://doi.org/10.1142/s2301385026500160

[19] Berghout, T., & Benbouzid, M. (2024). Fault Diagnosis in Drones via Multiverse Augmented Extreme Recurrent Expansion of Acoustic Emissions with Uncertainty Bayesian Optimisation. Machines, 12(8), 504. https://doi.org/10.3390/machines12080504

[20] Sonmezocak, T., & Yildiz, M. (2025). Hybrid Acoustic Fault Diagnosis in UAVs Using Wavelet Scattering Transform and Deep Learning. IEEE Access, 13, 159909-159919. https://doi.org/10.1109/access.2025.3607751

[21] Karkadakattil, A. (2026). Physics-Informed Telemetry Simulation and Shap-Based Explainable Machine Learning for Multi-Sensor UAV Fault Diagnosis. Transactions on Aerospace Research, 2026(2), 1-37. https://doi.org/10.2478/tar-2026-0007

[22] Ma, Y., Sun, Y., Chen, L., Wang, H., & Liu, B. (2025). Self-attention enhanced GraphSAGE for UAV fault diagnosis using vibration signals. Scientific Reports, 16(1), 2004. https://doi.org/10.1038/s41598-025-31632-8

[23] Chu, Z., Xing, S., Han, B., & Wang, J. (2024). A Novel Cross-Domain Mechanical Fault Diagnosis Method Fusing Acoustic and Vibration Signals by Vision Transformer. Sensors, 24(16), 5120. https://doi.org/10.3390/s24165120

[24] Huang, J., & Tong, J. (2026). Rolling Bearing Acoustic-Vibration Fusion Fault Diagnosis Based on Heterogeneous Modal Perception and Knowledge Distillation. Electronics, 15(8), 1631. https://doi.org/10.3390/electronics15081631

[25] Zhang, R., & Zhao, Z. (2026). A multimodal domain generalization bearing fault diagnosis method based on acoustic–vibration fusion. Measurement Science and Technology, 37(21), 216206. https://doi.org/10.1088/1361-6501/ae6de4

[26] Wang, S., Xu, Q., Zhang, K., Liu, Y., & Liu, H. (2025). Selection refines diagnosis: Mamba for acoustic weak fault diagnosis combining feature mode decomposition and selection. Advanced Engineering Informatics, 66, 103421. https://doi.org/10.1016/j.aei.2025.103421

[27] Liu, J., Hao, R., Liang, J., & Su, H. (2025). A gearbox fault diagnosis method based on CWT and Vision Mamba. Engineering Research Express, 7(3), 035524. https://doi.org/10.1088/2631-8695/adf0c4

[28] Chen, X., Chen, Y., Ju, Z., Qiang, Y., & Yang, J. (2026). Rolling bearing fault diagnosis under limited data using conditional mask-enhanced VQGAN and convolutional-embedded vision Mamba. Measurement Science and Technology, 37(4), 046206. https://doi.org/10.1088/1361-6501/ae362e

[29] Kang, Z., Yue, X., & Zhou, Y. (2026). SDMT-Net: a hybrid Spiral Dual Mamba and Transformer model for noise-robust rolling

bearing fault diagnosis. Measurement Science and Technology, 37(8), 086109. https://doi.org/10.1088/1361-6501/ae476e

[30] Cap, R. M., & Popa, C. A. (2025). Hybrid MambaVision and Transformer-Based Architecture for 3D Lane Detection. Sensors, 25(18), 5729. https://doi.org/10.3390/s25185729