Gearbox Fault Localisation Using Vibration and Acoustic Features with Implications for Maintenance Decisions

Shahrul Nazrin Khalid[1], Nor Azazi Ngatiman[1]*, Afdiman Anuar[1], Muhammad Noorazlan Shah Zainudin[2] and Sharif bin Mohd Zaki[1]

[1] Fakulti Teknologi dan Kejuruteraan Mekanikal, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia

[2] Faculty of Artificial Intelligence and Cyber Security, Universiti Teknikal Malaysia Melaka (UTeM), Melaka, Malaysia

DOI: https://doi.org/10.47772/IJRISS.2026.101100268

Received: 08 November 2026; Accepted: 13 November 2026; Published: 25 November 2026

ABSTRACT

Localising a gearbox fault to the responsible member of a gear pair can focus subsequent inspection, but pooled classification scores do not show how performance changes with operating speed or which component is selected when a prediction is wrong. This study compares vibration, airborne acoustic and hybrid descriptors for distinguishing a defective 25-tooth driving pinion (Condition B) from a defective 75-tooth driven gear (Condition C). A total of 50 records, each lasting five seconds, were collected at a sampling frequency of 6400 Hz across five operating speeds ranging from 400 to 2000 rpm. Five repetitions were obtained for each condition at every speed. Trial-level root-mean-square and Z-Freq features were classified using radial basis function support vector machines, K-nearest neighbours and bagged trees. Leave-one-speed-out validation used training-fold standardisation and deterministic seeds for stochastic models. Vibration KNN produced the highest pooled accuracy and macro-F1, 76.0% and 0.760. Its fold accuracy averaged 76.0 ± 25.1% and ranged from 50 to 100%. Hybrid KNN reached 68.0% and 0.678, while acoustic KNN reached 56.0% and 0.559. Vibration KNN made seven B-to-C and five C-to-B errors; vibration SVM made one and twelve. Fusion improved on acoustic inputs for all three classifiers, but exceeded vibration only for bagged trees. These results show that the preferred sensing and model combination depends on performance across operating speeds and on the direction of diagnostic error. The study provides a reproducible component-localisation comparison for this gearbox and a basis for broader validation with healthy states, additional loads, defect severities and independent gearboxes. The workflow provides a transparent basis for future validation under broader gearbox operating conditions.

Keywords: gear fault localisation, condition-based maintenance, vibration and acoustics, sensor fusion, machine learning

INTRODUCTION

A maintenance team investigating a gearbox needs to determine which component warrants inspection. Confusing a faulty driving pinion with its mating driven gear may direct attention to the wrong member. Condition-based maintenance connects measured equipment condition with decisions about maintenance activities (Jardine et al., 2006). A diagnostic model is useful in that process only when its errors and operating limits are understood.

This study examines fault localisation between two defective gear states. It compares compact vibration, acoustic and hybrid descriptors across speed, then interprets classification errors as potential errors in directing component inspection. The evidence concerns the measured test rig and supports an exploratory discussion of maintenance decisions. It provides no measurement of repair cost, inspection duration or production downtime.

LITERATURE REVIEW

Condition monitoring and maintenance decisions

Jardine et al. (2006) distinguish acquisition, processing and decision-making within condition-based maintenance. A classification score describes diagnostic behaviour, whereas an intervention also depends on the consequences of error. Lei et al. (2020) relate fault-diagnosis models to practical operating conditions. In another manufacturing application, Kasim et al. (2021) improved the repeatability of tool-wear region recognition by controlling initial clustering centroids. Although the task differs, it shows why stochastic settings should be documented. Here, predictions provide evidence for inspection rather than automatic replacement instructions.

Vibration and acoustic information

An accelerometer measures structural response; a microphone captures pressure radiated into air. Lu et al. (2021) investigated airborne sound during progressive helical-gearbox deterioration, and Yao et al. (2018) studied sound-based gear diagnosis. Li et al. (2016) fused an acoustic-emission sensor with vibration, using a sensing mechanism different from the microphone in this study. These studies motivate sensor comparisons but do not establish that fusion will improve every gearbox dataset.

Feature fusion and validation

RMS summarises amplitude, while Z-Freq combines dispersion and kurtosis of spectral amplitudes (Ngatiman et al., 2018). Othman et al. (2024) extended Z-Freq to two sensors. Multimodal performance depends on representation and fusion design (Baltrušaitis et al., 2019; Tsanousa et al., 2022). Abdul et al. (2016) examined speed- and load-held-out gear classification, while Kapoor and Narayanan (2023) show how leakage between related records can inflate performance. Withholding one complete speed therefore tests transfer to an unseen operating condition (Roberts et al., 2017).

METHODOLOGY

Experimental design and data records

The gear pair comprised a 25-tooth driving pinion and a 75-tooth driven gear, giving a nominal 3:1 ratio. Table 1 defines the two physical states. Defect severity is not assumed equivalent between the specimens. Five recordings per condition at 400, 800, 1200, 1600 and 2000 rpm produced 50 trials, with 25 observations in each class.

Table 1: Physical gear states used for binary fault localisation.

Condition	Driving pinion	Driven gear	Trials
B	Defective	Healthy	25
	25 teeth	75 teeth	
C	Healthy	Defective	25
	25 teeth	75 teeth	

A PCB Piezotronics 352C33 casing accelerometer and GRAS 40PP-10 free-field microphone were acquired synchronously using NI FlexLogger. Each five-second record contained 32,000 samples at 6400 Hz. Acceleration and pressure were expressed in m/s² and Pa. Speed was read from the rig display; no tachometer signal was used.

Signal processing and predictor construction

The processing routine uses nine mean-centred, one-second windows per record, each containing N = 6400

samples. A 3200-sample step gives 50% overlap. Window RMS is defined by Eq. (1), where x denotes centred samples. Each descriptor is averaged over its nine windows to produce one trial-level value.

$$RMS_w = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_{w,i}^{\,2}} \quad (1)$$

For Z-Freq, a symmetric Hann window precedes the single-sided FFT with coherent-gain correction. Interior amplitudes are doubled, while Nyquist remains undoubled. Removing DC retains Nf = 3200 positive-frequency bins at 1 Hz spacing. Affix and Annex regions cover $0 < f \leq 1600$ Hz and $1600 < f \leq 3200$ Hz. In Eq. (2), s is population standard deviation and κ is Pearson kurtosis of spectral amplitudes within each region (Ngatiman et al., 2018).

$$Z_w = \frac{1}{N_f}\sqrt{\kappa_{Aff} s_{Aff}^{\,4} + \kappa_{Ann} s_{Ann}^{\,4}} \quad (2)$$

Single-channel Z-Freq has units m^2/s^4 for acceleration and Pa^2 for pressure. The vibration input contains vibration RMS and Z-Freq; the acoustic input contains acoustic RMS and Z-Freq. The hybrid input contains vibration RMS, acoustic RMS and H_ZFreq. Its construction adapts the two-sensor formulation of Othman et al. (2024):

$$H_{Z,w} = \frac{1}{N_f}\sqrt{K_{Vib,w} S_{Vib,w}^{\,4} + K_{Ac,w} S_{Ac,w}^{\,4}} \quad (3)$$

For each synchronous window, S and K are the population standard deviation and Pearson kurtosis of each sensor's 3200 matched positive-frequency amplitudes. The spectra retain their measured units during construction of H_ZFreq, so the mixed-unit coefficient can change with sensor calibration or amplitude scaling. Training-fold standardisation controls the predictor scale presented to the classifier but does not remove that sensitivity from H_ZFreq. Equation (3) is evaluated per window and then averaged within the trial; it is not reconstructed from trial-averaged moments or the two single-channel coefficients. Sensor-scale sensitivity was not varied experimentally.

Classifiers and speed-held-out evaluation

Models were fitted in MATLAB R2024b Update 9 using the Statistics and Machine Learning Toolbox. RBF-SVM uses automatic kernel scale and box constraint 1 through one-versus-one error-correcting output codes. KNN uses five neighbours, Euclidean distance and squared-inverse-distance weighting. Bagging uses 200 trees, minimum leaf size 2 and a maximum of 30 splits. The base random seed was 2026, with deterministic modality, classifier and fold offsets. Predictor sets and settings were fixed, with no additional feature selection.

Leave-one-speed-out validation holds out all ten trials from one speed and trains on the other 40. The same five folds are used for every model. Means and sample standard deviations estimated from training records are applied to test records; zero or nonfinite scales are replaced by one. Accuracy and macro-F1 are calculated for pooled predictions and each fold. Fold results are reported as mean, sample standard deviation, range and a descriptive two-sided 95% t interval. Because there are only five folds, these intervals describe variation between speeds rather than uncertainty across independent gearboxes (Varoquaux, 2018).

Trial-level feature exports were joined by condition, speed and trial identifier; their vibration RMS and Z-Freq fields agreed exactly. The workflow was rerun from all 50 feature vectors, producing 45 fold summaries and 450 trial predictions. Confusion counts and summary metrics were reconciled. The rerun reproduces the machine-learning stage from archived features; extraction was not repeated from raw recordings.

RESULTS

Amplitude response across operating speeds

Table 2 and Figure 1 summarise RMS using five trial values per condition and speed. The pooled vibration means were 0.9419 m/s² for B and 0.8018 m/s² for C. Corresponding acoustic means were 0.2564 and 0.2411 Pa. These averages conceal speed-dependent changes in class ordering.

Table 2: RMS by driving speed and condition, reported as mean ± sample standard deviation of five trials.

Speed(rpm)	Condition	Vibration RMS(m/s²)	Acoustic RMS (Pa)
400	B	0.3717 ± 0.0407	0.1249 ± 0.0056
400	C	0.2747 ± 0.0076	0.0864 ± 0.0018
800	B	0.6486 ± 0.0164	0.2022 ± 0.0069
800	C	0.7649 ± 0.0086	0.1512 ± 0.0024
1200	B	1.1101 ± 0.0605	0.2825 ± 0.0065
1200	C	0.8040 ± 0.0165	0.1795 ± 0.0039
1600	B	1.4846 ± 0.1182	0.2944 ± 0.0104
1600	C	1.0065 ± 0.0124	0.2482 ± 0.0020
2000	B	1.0947 ± 0.1543	0.3781 ± 0.0059
2000	C	1.1590 ± 0.0288	0.5401 ± 0.0161

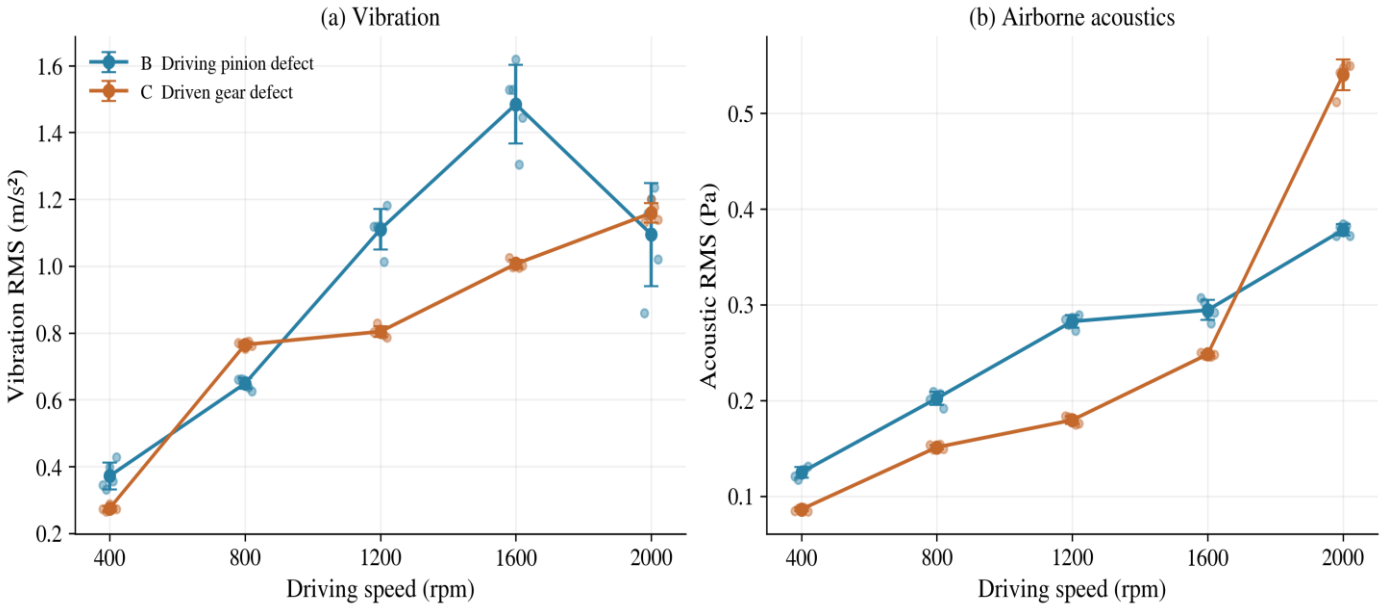

Figure 1: Trial-level RMS for vibration and airborne acoustics. Small points show individual trials; connected markers and error bars represent mean ± sample standard deviation.

Vibration RMS was higher for B at 400, 1200 and 1600 rpm, and for C at 800 and 2000 rpm. B peaked at 1.4846 m/s² at 1600 rpm, whereas C increased throughout. Acoustic RMS was higher for B through 1600 rpm; at 2000 rpm, C reached 0.5401 Pa and exceeded B at 0.3781 Pa. The changes in class ordering show why amplitude thresholds must account for operating speed. They do not establish equivalent defect severity.

Spectral statistical response

Vibration Z-Freq was higher for B at all five speeds (Table 3 and Figure 2). Its pooled means were 5.413 × 10^{-6} m^2/s^4 for B and 1.458 × 10^{-6} m^2/s^4 for C. B peaked at 1600 rpm and then declined. This consistent ordering differs from the RMS ordering at 800 and 2000 rpm.

Table 3: Z-Freq by speed and condition, reported as mean ± sample standard deviation of five trials. Both numerical columns use the multiplier 10^{-6}.

Speed (rpm)	Condition	Vibration Z-Freq (× 10^{-6} m^2/s^4)	Acoustic Z-Freq (× 10^{-6} Pa^2)
400	B	0.6092 ± 0.2231	0.0389 ± 0.0057
400	C	0.0887 ± 0.0018	0.0312 ± 0.0021
800	B	1.3751 ± 0.2052	0.1037 ± 0.0196
800	C	0.7414 ± 0.0629	0.0264 ± 0.0030
1200	B	7.8816 ± 0.4395	0.2112 ± 0.0187
1200	C	0.6372 ± 0.0424	0.0444 ± 0.0058
1600	B	12.2426 ± 2.9158	0.1805 ± 0.0277
1600	C	2.6003 ± 0.0650	0.2462 ± 0.0067
2000	B	4.9569 ± 2.4334	0.2746 ± 0.0108
2000	C	3.2244 ± 0.1301	1.8236 ± 0.1056

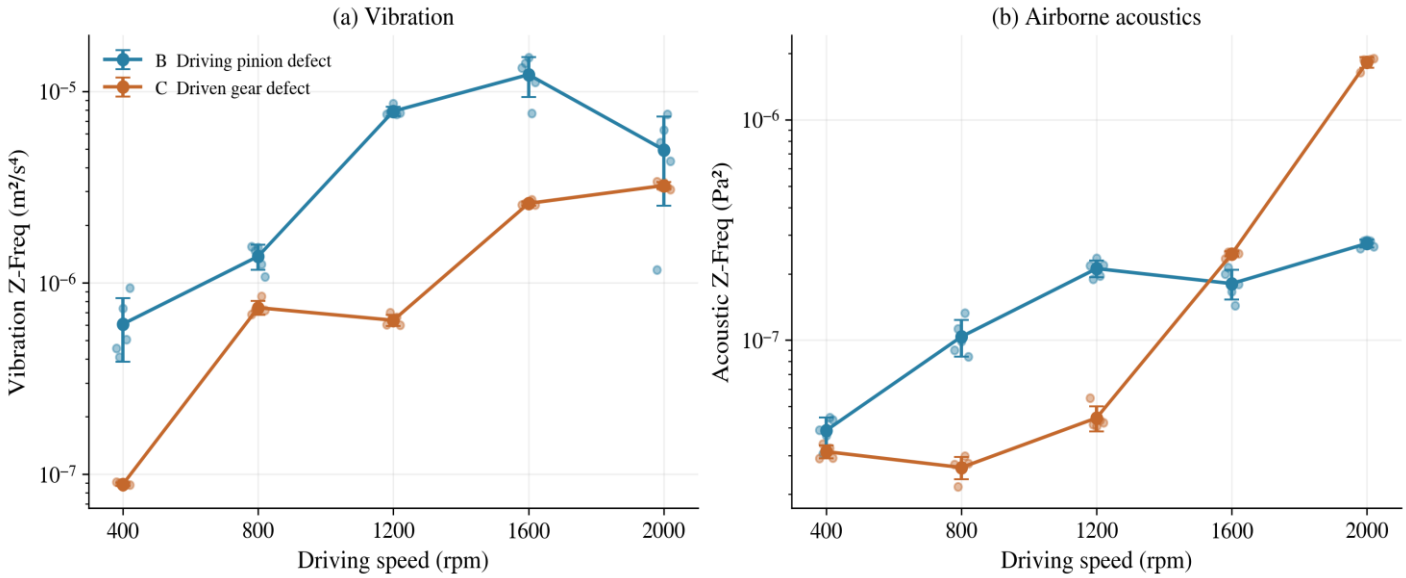

Figure 2: Trial-level Z-Freq for vibration and acoustic pressure on logarithmic vertical axes. Points show individual trials; error bars represent mean ± sample standard deviation.

Acoustic Z-Freq was higher for B at 400, 800 and 1200 rpm, and for C at 1600 and 2000 rpm. At 2000 rpm, the means were 2.746 × 10^{-7} Pa^2 for B and 1.824 × 10^{-6} Pa^2 for C. At 1600 rpm, C had the larger acoustic coefficient despite lower acoustic RMS, indicating a difference in spectral-amplitude distribution. Consistent descriptive ordering of vibration Z-Freq does not establish reliable classification at a withheld speed; that question is assessed separately through LOSO.

Classification across unseen operating speeds

Vibration KNN correctly classified 38 of the 50 trials, giving the highest pooled accuracy of 76.0% and macro-F1 of 0.760 (Table 4). Hybrid KNN achieved 68.0% and 0.678, while acoustic KNN achieved 56.0% and

0.559. The best hybrid model classified four fewer trials correctly than vibration KNN and six more than acoustic KNN.

Table 4: Pooled LOSO performance and variation across the five held-out speeds. Fold values are mean ± sample standard deviation.

Input	Classifier	Pooled Acc. / F1	Fold Acc. mean ± SD (%)	Fold macro-F1 mean ± SD
Vibration	RBF SVM	74.0 / 0.727	74.0 ± 25.1	0.673 ± 0.334
Vibration	KNN	76.0 / 0.760	76.0 ± 25.1	0.692 ± 0.338
Vibration	Bagged trees	56.0 / 0.548	56.0 ± 26.1	0.446 ± 0.313
Acoustic	RBF SVM	52.0 / 0.517	52.0 ± 24.9	0.413 ± 0.281
Acoustic	KNN	56.0 / 0.559	56.0 ± 32.9	0.486 ± 0.354
Acoustic	Bagged trees	42.0 / 0.408	42.0 ± 13.0	0.290 ± 0.072
Hybrid	RBF SVM	62.0 / 0.601	62.0 ± 21.7	0.505 ± 0.289
Hybrid	KNN	68.0 / 0.678	68.0 ± 29.5	0.590 ± 0.374
Hybrid	Bagged trees	58.0 / 0.580	58.0 ± 23.9	0.457 ± 0.304

Hybrid accuracy was lower than vibration for SVM (62.0% versus 74.0%) and KNN (68.0% versus 76.0%), but slightly higher for bagged trees (58.0% versus 56.0%). Relative to acoustic inputs, hybrid inputs improved accuracy by 10, 12 and 16 percentage points for SVM, KNN and bagging, respectively. Fusion therefore improved on the acoustic-only input in every classifier family, although its advantage over vibration depended on the classifier.

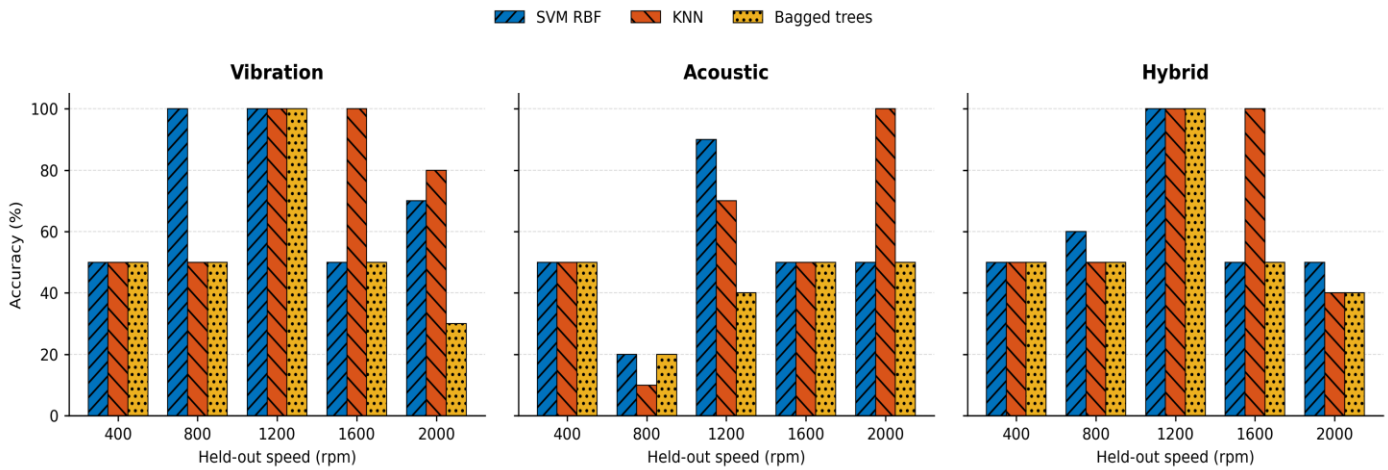

Figure 3: Fold-wise classification accuracy when each operating speed was held out in turn. Each bar represents the accuracy obtained for the ten trials at the indicated held-out speed; Table 4 reports the pooled metrics and fold-wise mean ± sample standard deviation.

Performance varied substantially between held-out speeds. Vibration KNN accuracy ranged from 50-100%, hybrid KNN from 40-100%, and acoustic KNN from 10-100%. Their descriptive 95% t intervals for mean fold accuracy were 44.8-100.0%, 31.4-100.0% and 15.2-96.8%, respectively. These broad intervals show why pooled and speed-specific results should be considered together.

Class-specific errors and potential inspection direction

Figure 4 shows all nine confusion matrices, with observed classes in rows and predictions in columns. A B-to-C error would direct inspection towards the driven gear when the recorded defect was in the driving pinion. A

C-to-B error would have the reverse implication. These are interpretations of classification counts; inspection outcomes were not observed.

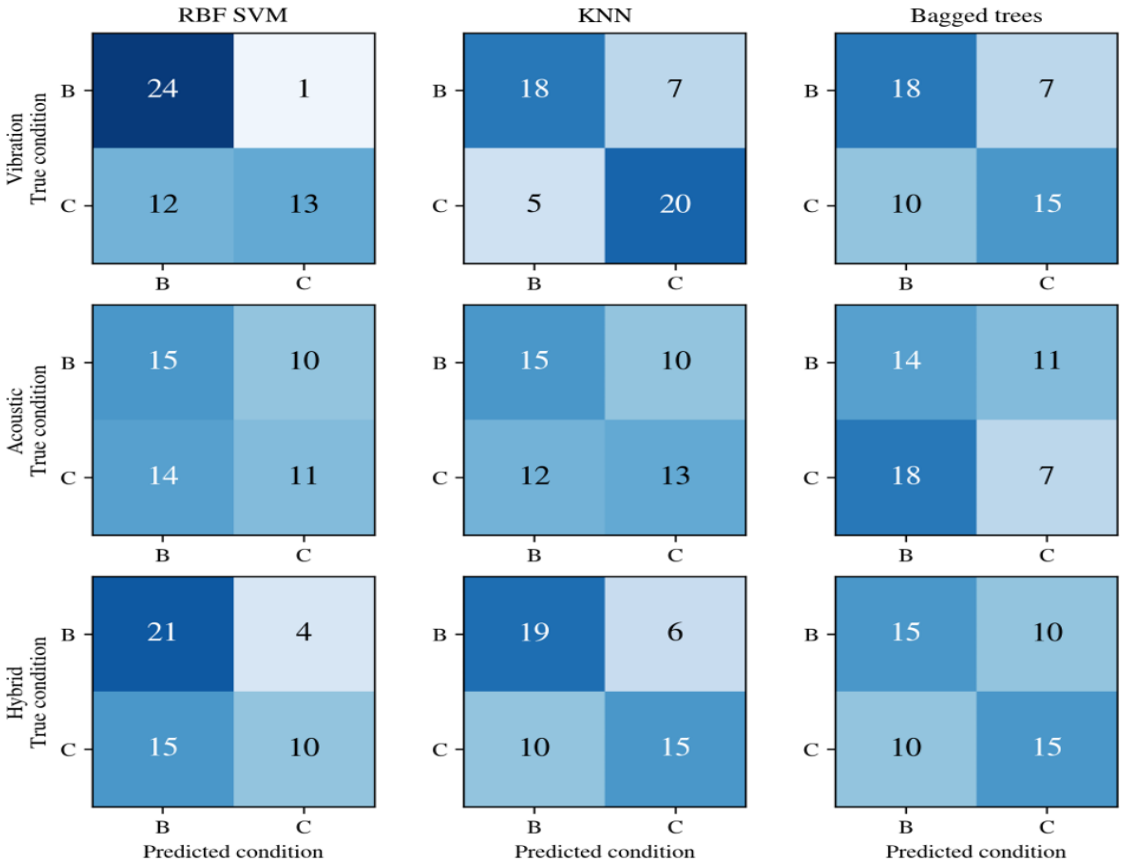

Figure 4: Pooled LOSO confusion matrices from the deterministic rerun. Panel rows denote vibration, acoustic and hybrid inputs; columns denote RBF SVM, KNN and bagged trees. Rows within each panel are observed classes B/C and columns are predicted classes B/C.

Vibration KNN correctly identified 18 B trials and 20 C trials, leaving seven B-to-C and five C-to-B errors. Hybrid KNN identified 19 B and 15 C trials. It added one correct B classification but lost five correct C classifications, explaining its four-trial reduction in accuracy. Vibration SVM produced only one B-to-C error but twelve C-to-B errors; its recalls were 96.0% for B and 52.0% for C. The preferred error balance therefore depends on the consequence assigned to each direction.

DISCUSSION

Interpreting features for component inspection

The observed feature patterns caution against assigning inspection priority from amplitude alone. RMS ordering changes with speed, and a large coefficient does not establish greater damage severity. Although vibration Z-Freq remained higher for B, the experiment contains one predefined defect state per component. Specimen effects and transmission paths may contribute to separation. A diagnostic label should consequently be interpreted alongside the operating speed and evidence from physical inspection.

Illustrative consequences of diagnostic error

Following the mean misclassification cost framework (Komisarenko & Kull, 2025), Eq. (4) weights B-to-C errors by ρ relative to C-to-B errors. The illustration assumes the experimental class balance, constant positive consequences within each error direction and zero loss for correct classifications:

$$R_m(\rho) = \frac{\rho\, n_{B\to C,m} + n_{C\to B,m}}{50} \quad (4)$$

Here, n denotes a pooled error count. Table 5 compares all nine configurations at assumed ratios ρ = 1 and ρ = 2. The ratios are illustrative scenarios rather than estimates from maintenance records. The resulting values are relative losses, not measured costs or error percentages.

Table 5: Illustrative relative loss for all nine LOSO configurations under assumed consequence ratios.

Input	Classifier	B → C errors	C → B errors	Loss ρ = 1	Loss ρ = 2
Vibration	RBF SVM	1	12	0.26	0.28
Vibration	KNN	7	5	0.24	0.38
Vibration	Bagged trees	7	15	0.44	0.58
Acoustic	RBF SVM	10	14	0.48	0.68
Acoustic	KNN	10	12	0.44	0.64
Acoustic	Bagged trees	11	18	0.58	0.8
Hybrid	RBF SVM	4	15	0.38	0.46
Hybrid	KNN	6	10	0.32	0.44
Hybrid	Bagged trees	11	10	0.42	0.64

Vibration KNN has the lowest loss at ρ = 1 (0.24), and vibration SVM at ρ = 2 (0.28). Across all nine configurations, KNN minimises loss for $0 < \rho < 7/6$ and SVM for $\rho > 7/6$; they tie where $7\rho + 5 = \rho + 12$. This post hoc calculation is not an independently validated model-selection rule.

Sensor choice and the value of fusion

The hybrid input showed no consistent advantage over vibration. It performed better for bagged trees but worse for SVM and KNN. The mixed-unit H_ZFreq coefficient retains the original sensor amplitude scales and combines two spectral distributions in one value, so calibration and representation choices can affect the information presented to a classifier. These effects were not separated in the present experiment. Fusion should therefore be assessed through its measured contribution for a chosen classifier and through the acquisition requirements of the additional sensor (Baltrušaitis et al., 2019; Tsanousa et al., 2022).

Practical relevance and scope of the findings

The findings provide a practical basis for identifying which member of the gear pair should receive priority during component-level inspection. Among the evaluated configurations, vibration KNN achieved the highest overall classification performance, while the directional-error analysis showed how models with similar accuracy can lead to different inspection priorities. Considering whether an error directs attention towards the driving pinion or the driven gear therefore adds information that is directly relevant to maintenance planning.

The balanced representation of Conditions B and C enabled an equitable comparison of the classifiers and sensing inputs. It also allowed class-specific recall and error direction to be examined without either condition dominating the results. For industrial use, the relative-loss calculation can be adapted using site-specific estimates of fault prevalence, inspection effort and the consequence of examining the incorrect component. This provides a transparent route for combining diagnostic performance with local maintenance priorities.

The present task focuses on localising a defect between two predefined faulty gear states. Within this scope, a B or C prediction identifies the component that warrants closer examination and can support more focused inspection. Repair or replacement decisions would continue to draw on physical confirmation and the operating context. Adding healthy, progressive-defect and variable-load records in future work would extend the framework towards fault detection, damage development and intervention timing.

The evaluation used 50 trials from one experimental rig at five steady operating speeds. Leave-one-speed-out validation provided a structured test of transfer to an operating speed excluded from training, while the fold summaries identified the speeds at which performance was least stable. A next-stage study can build directly on this protocol by adding independent gearboxes, specimens, loads, measurement sessions and graded defect severity. Healthy records would extend the current localisation task to fault detection followed by localisation. Retaining trial-level predictions, the deterministic MATLAB settings and the archived feature data provides a reproducible baseline for those extensions.

CONCLUSION

This study examined vibration, airborne acoustic and hybrid features for distinguishing a defective driving pinion from a defective driven gear. Vibration Z-Freq was higher for B throughout the five speeds, whereas RMS and acoustic Z-Freq changed class ordering. Vibration KNN achieved the highest observed LOSO accuracy of 76.0% and macro-F1 of 0.760. Hybrid KNN reached 68.0% and 0.678, and acoustic KNN reached 56.0% and 0.559.

The maintenance implication lies in the direction of diagnostic error. Vibration KNN produced fewer total errors than vibration SVM, while SVM produced fewer errors that would direct inspection away from a defective pinion. Under the stated illustrative assumptions, their relative preference changes at a consequence ratio of 7/6. The fold-wise analysis also shows that performance depends on the speed withheld from training. Hybrid features improved on acoustic features for all three classifiers, but their comparison with vibration depended on classifier choice. The deterministic workflow and fold-level outputs provide a clear baseline for validation with measured maintenance consequences and independent gearboxes.

ACKNOWLEDGEMENT

This research was supported by Universiti Teknikal Malaysia Melaka (UTeM) through the Short-Term Research Grant (Grant No. PJP/2025/FTKM/PERSPEKTIF/SA0119). The authors gratefully acknowledge the Faculty of Mechanical Technology and Engineering (FTKM), UTeM, for providing the laboratory facilities and technical support required to conduct this study.

DATA AVAILABILITY

The trial-level feature dataset, fold-wise predictions, pooled confusion matrices, model settings and MATLAB code supporting the reported analyses are available from the corresponding author upon reasonable request.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1. Abdul, Z. K., Al-Talabani, A., & Abdulrahman, A. O. (2016). A new feature extraction technique based on 1D local binary pattern for gear fault detection. *Shock and Vibration, 2016*, Article 8538165. https://doi.org/10.1155/2016/8538165
2. Baltrušaitis, T., Ahuja, C., & Morency, L.-P. (2019). Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 41*(2), 423-443. https://doi.org/10.1109/TPAMI.2018.2798607
3. Jardine, A. K. S., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing, 20*(7), 1483-1510. https://doi.org/10.1016/j.ymssp.2005.09.012
4. Kapoor, S., & Narayanan, A. (2023). Leakage and the reproducibility crisis in machine-learning-based science. *Patterns, 4*(9), Article 100804. https://doi.org/10.1016/j.patter.2023.100804

5. Kasim, N. A., Nuawi, M. Z., Ghani, J. A., Rizal, M., Ngatiman, N. A., & Haron, C. H. C. (2021). Enhancing clustering algorithm with initial centroids in tool wear region recognition. *International Journal of Precision Engineering and Manufacturing, 22*(5), 843-863. https://doi.org/10.1007/s12541-020-00450-5
6. Komisarenko, V., & Kull, M. (2025). Cost-sensitive classification with cost uncertainty: Do we need surrogate losses? *Machine Learning, 114*, Article 132. https://doi.org/10.1007/s10994-024-06634-8
7. Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., & Nandi, A. K. (2020). Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mechanical Systems and Signal Processing, 138*, Article 106587. https://doi.org/10.1016/j.ymssp.2019.106587
8. Li, C., Sanchez, R.-V., Zurita, G., Cerrada, M., Cabrera, D., & Vásquez, R. E. (2016). Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signals. *Mechanical Systems and Signal Processing, 76-77*, 283-293. https://doi.org/10.1016/j.ymssp.2016.02.007
9. Lu, K., Gu, J. X., Fan, H., Sun, X., Li, B., & Gu, F. (2021). Acoustics based monitoring and diagnostics for the progressive deterioration of helical gearboxes. *Chinese Journal of Mechanical Engineering, 34*, Article 82. https://doi.org/10.1186/s10033-021-00603-1
10. Ngatiman, N. A., Nuawi, M. Z., & Abdullah, S. (2018). Z-Freq: Signal analysis-based gasoline engine monitoring technique using piezo-film sensor. *International Journal of Mechanical Engineering and Technology, 9*(5), 897-910. https://iaeme.com/Home/article_id/IJMET_09_05_099
11. Othman, M. N. B., Nuawi, M. Z., Ngatiman, N. A., & Yusri, M. Y. (2024). Z-Freq hybrid: Signal analysis based car air conditioning compressor monitoring technique using accelerometer and piezo-film sensor. *Journal of Physics: Conference Series, 2721*(1), Article 012021. https://doi.org/10.1088/1742-6596/2721/1/012021
12. Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography, 40*, 913-929. https://doi.org/10.1111/ecog.02881
13. Tsanousa, A., Bektsis, E., Kyriakopoulos, C., Gómez González, A., Leturiondo, U., Gialampoukidis, I., Karakostas, A., Vrochidis, S., & Kompatsiaris, I. (2022). A review of multisensor data fusion solutions in smart manufacturing: Systems and trends. *Sensors, 22*(5), Article 1734. https://doi.org/10.3390/s22051734
14. Varoquaux, G. (2018). Cross-validation failure: Small sample sizes lead to large error bars. *NeuroImage, 180*, 68-77. https://doi.org/10.1016/j.neuroimage.2017.06.061
15. Yao, Y., Wang, H., Li, S., Liu, Z., Gui, G., Dan, Y., & Hu, J. (2018). End-to-end convolutional neural network model for gear fault diagnosis based on sound signals. *Applied Sciences, 8*(9), Article 1584. https://doi.org/10.3390/app8091584