

Received: 23 February 2026 / Accepted: 20 March 2026 / Published online: 28 April 2026

*gearbox fault diagnosis,
vibration signal analysis,
hybrid machine learning,
predictive maintenance*

Bakhtiyar MAMMADLI¹, Jakub RZECZKOWSKI²,
Lukasz MAZUREK³, Jarosław JASINSKI¹,
Karol FRYDRYCH¹, Azad MAMMADLI⁴,
F. Javier DOMINGUEZ-GUTIERREZ^{*1}

HYBRID MACHINE LEARNING FRAMEWORK FOR DIRECTION-DEPENDENT GEARBOX FAULT DETECTION

Reliable quality control and fault diagnosis are essential for ensuring machine reliability and preventing unexpected failures. One of the critical machine components for which such a diagnosis enables failure-free, long-term exploitation is gearboxes. Conventional vibration-based monitoring often depends on expert interpretation of signal patterns and gear-mesh behaviour, which limits scalability and consistency. In this work, a hybrid machine-learning framework for binary gearbox health classification using engineered vibration features. Time and frequency domain descriptors capturing impulsiveness, gear-mesh spectral characteristics, and modulation effects were extracted from tri-axial acceleration signals. To account for direction-dependent dynamics, separate models were developed for left (RPM0) and right (RPM1) rotational conditions. We employ an unsupervised Isolation Forest trained exclusively on healthy data for anomaly detection, and a supervised Logistic Regression classifier trained on both healthy and faulty samples. Predefined decision thresholds were applied to ensure methodological transparency and minimize overfitting. Evaluation on independent test cases demonstrates that direction-specific modelling combined with physically interpretable features enables robust gearbox fault detection. The proposed framework provides a reproducible and industrially applicable strategy for automated condition monitoring. Such an approach will provide precise solutions for early fault detection, predictive maintenance scheduling, and real-time performance optimization of gearboxes and machinery systems.

1. INTRODUCTION

Gearboxes are critical components in a wide range of industrial systems, including manufacturing equipment, wind turbines, heavy machinery and robotic systems. Akmease and Eastham [1] investigated design aspects of electromechanical drive systems, highlighting

¹ NOMATEN Center of Excellence, National Center for Nuclear Research, Poland

² Faculty of Mechanical Engineering, Lublin University of Technology, Poland

³ ABM Greiffenberger Polska Sp. z o.o, Poland

⁴ Batabat Astrophysical Observatory, Nakhchivan State University, Azerbaijan

* E-mail: javier.dominguez@ncbj.gov.pl

<https://doi.org/10.36897/jme/219831>

the importance of reliable transmission mechanisms in rotating machinery. Yaqoob et al. [2] proposed a gear selection framework aimed at improving accuracy in industrial robotic systems, demonstrating the critical role of gearbox performance in precision applications. Dinh et al. [3] studied multi-objective optimization of a two-stage helical gearbox, illustrating the engineering complexity associated with gearbox design and performance optimization. Their continuous, reliable operation is essential to maintaining high system availability and preventing unplanned downtime, which may result in substantial economic losses and safety risks. Because gearboxes are frequently subjected to variable loading conditions, harsh operating environments, and extended service periods, early fault detection and accurate defect localization are fundamental requirements for effective predictive maintenance strategies [4]. Vibration-based condition monitoring has long been established as the primary diagnostic technique for gearbox health assessment. Conventional approaches rely on the analysis of characteristic frequency components associated with gear-mesh frequencies, sidebands and harmonics, combined with expert interpretation of time- and frequency-domain signal patterns. These techniques are grounded in the physical understanding of gearbox dynamics and failure mechanisms [5]. However, they typically require considerable domain expertise, manual threshold definition, and machine-specific tuning. Such requirements limit scalability and robustness, particularly in large industrial fleets where operating regimes vary and fault data may be sparse or incomplete.

In recent years, machine learning (ML) techniques have emerged as powerful alternatives to traditional rule-based diagnostics, enabling automated extraction of discriminative features and improved adaptability to complex signal behavior. Supervised machine-learning methods have been widely applied to engineering diagnostics and prediction tasks. Tuyet-Doan et al. [6] demonstrated the effectiveness of the XGBoost algorithm for gearbox fault diagnosis using vibration signals from road-vehicle transmissions. Mammadli et al. [7] applied data-driven machine-learning techniques to predict the axial load capacity of steel structural components, illustrating the capability of ML models to capture complex relationships in engineering datasets. In a related context, Rzeczkowski [8] employed deep-learning techniques for acoustic-emission signal analysis to characterize delamination processes in composite laminates, highlighting the growing role of advanced ML methods in signal-based damage detection. More recently, deep learning architectures have been explored for end-to-end fault diagnosis, allowing automatic feature learning directly from raw vibration signals [9, 4]. Although these approaches can achieve high predictive accuracy, they often require large, well-annotated datasets and substantial computational resources, which may restrict their applicability in real-time industrial environments [10].

A key practical challenge in industrial condition monitoring is the scarcity and imbalance of fault data. In typical operating scenarios, gearboxes remain in healthy conditions for the majority of their service life, while fault occurrences are rare, diverse, and not always systematically recorded [11]. This strong class imbalance motivates the use of unsupervised or semi-supervised ML techniques that model normal operating behaviour and identify deviations without relying on explicit fault labels. Anomaly detection algorithms such as Isolation Forest, one-class Support Vector Machines, and autoencoder-based methods have therefore attracted increasing attention as viable solutions for early fault detection under

realistic industrial constraints. From an application standpoint, effective gearbox diagnostics should not only detect abnormal operation but also provide actionable information regarding the likely fault source. Consequently, machine-learning frameworks that combine anomaly detection and fault classification have been increasingly explored for industrial diagnostics. Primera et al. [12] developed an IoT-based diagnostic analytics system integrating machine learning for the evaluation of heavy-industry gearbox faults. Nguyen and Diep [13] proposed a deep-learning approach based on a multi-scale convolutional neural network for vibration-based gearbox fault diagnosis. In addition, Suryawanshi et al. [14] provided a comprehensive review of machine-learning and deep-learning methods for gearbox diagnostics, highlighting the growing interest in hybrid and data-driven monitoring strategies. In such frameworks, an unsupervised model first detects deviations from healthy behaviour, and a supervised classifier subsequently identifies the specific fault type or defective component. This modular structure aligns well with industrial maintenance workflows, enabling reliable anomaly detection even when fault libraries are incomplete, while allowing more detailed classification once labelled fault data becomes available [15]. In addition to predictive performance, industrial deployment requires ML solutions to be interpretable, computationally efficient, and compatible with real-time implementation. Feature-based approaches that combine physically informed signal descriptors with robust classification algorithms remain particularly attractive in this context. By incorporating domain knowledge, such as features extracted around gear-mesh frequencies and their harmonics, these methods preserve physical interpretability while leveraging the pattern-recognition capability of modern ML techniques [16, 17].

In this manuscript, we propose a novel hybrid diagnostic framework combining unsupervised anomaly detection and supervised classification is proposed for gearbox fault detection under realistic industrial constraints. Physically informed vibration features are extracted in both time and frequency domains, including mesh-harmonic-based descriptors and statistical indicators. An Isolation Forest model is employed to learn normal operating behaviour using healthy data only, while a Logistic Regression classifier is trained to distinguish between healthy and faulty conditions when labelled data are available. Separate models are constructed for different rotation directions to account for operational asymmetries. The framework is designed to ensure robustness, transparency and reproducibility, while avoiding extensive hyperparameter tuning. The effectiveness and practical implications of the proposed approach are analysed using experimentally obtained vibration datasets, with particular emphasis on balanced fault detection performance and interpretability for industrial application.

2. METHODS

2.1 EXPERIMENTAL METHODS

2.1.1. EXPERIMENTAL SETUP

The gearbox diagnostic test rig was configured as a semi-automated station integrated within a production environment. The test rig consisted of two principal subsystems:

a mechanical test chamber and an integrated control and measurement system. The setup was designed to accommodate gearboxes of different sizes within the same product family and to ensure repeatable and standardized testing conditions. The gearboxes under investigation were mounted on a dedicated trolley providing precise alignment and stable fixation during operation as presented in Fig. 1. The chamber was manually closed prior to test initiation and opened after completion. Its status was automatically supervised by the control system and test execution was enabled only when proper closure was confirmed, ensuring operational safety and measurement reliability.

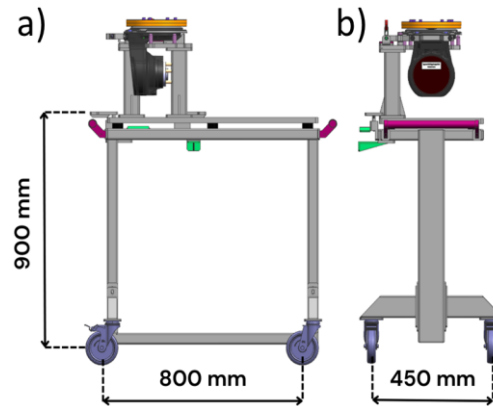


Fig. 1. Test rig with the gearbox mounted on the dedicated assembly trolley: (a) side view showing the overall structure and main dimensions (height 900 mm, length 800 mm); (b) front view showing the trolley width (450 mm)

Vibration signals were acquired using a Hansford Sensors HS173 industrial accelerometer. The sensor was a triaxial IEPE accelerometer enabling simultaneous measurement in three orthogonal directions. The accelerometer was characterized by a nominal sensitivity of 100 mV/g, a measurement range of ± 50 g and a frequency response suitable for gearbox diagnostics (up to approximately 10 kHz ± 3 dB). The sensor was mounted directly to the gearbox housing using a specially designed measurement flange to ensure repeatable positioning and high structural coupling stiffness as shown in Fig. 2.

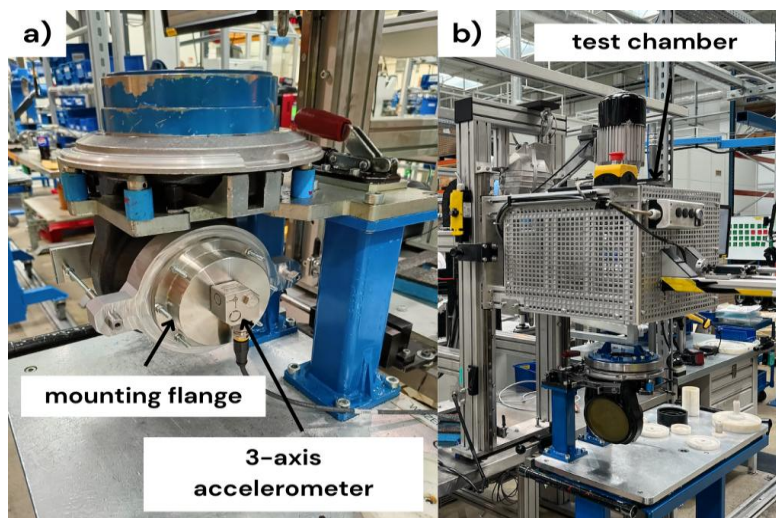


Fig. 2. Test stand: a) gearbox equipped with the dedicated measurement flange used for accelerometer mounting providing stable and repeatable vibration measurements, b) the test chamber enclosing the measurement area

The vibration signals were recorded using a National Instruments NI 9234 dynamic signal acquisition module. The NI 9234 module provided four simultaneously sampled analog input channels with 24-bit resolution and integrated IEPE signal conditioning. The module incorporated built-in anti-aliasing filters with automatic cutoff adjustment dependent on the selected sampling frequency ensuring proper band limitation prior to analog-to-digital conversion. Data were sampled at a frequency of 12 800 Hz. Measurements were conducted at a constant rotational speed of 3000 rpm. The duration of a single gearbox test was 2 minutes. Vibration signals were recorded in two operational directions denoted as RPM0 and RPM1 corresponding to left and right rotational directions respectively. The acquired data were stored for further post-processing and diagnostic feature extraction.

2.1.2 SIGNAL PROCESSING AND FEATURE ENGINEERING

The vibration signals were processed using a physics-informed feature engineering framework tailored to gear-mesh-based fault detection. Each measurement consisted of segmented time-series data structured as $N \times 3 \times L$, where N denotes one-second windows, three channels correspond to orthogonal accelerometer axes (X, Y, Z), and $L = 12800$ samples were recorded at a sampling frequency of $f_s = 12.8$ kHz. All signals were mean-centered prior to feature extraction.

The shaft rotational frequency was computed as

$$f_{rot} = \frac{n_{rpm}}{60}, \quad (1)$$

and theoretical gear-mesh frequencies were derived based on gearbox kinematics. The primary mesh frequency was defined as

$$f_{m12} = z_1 f_{rot} \quad (2)$$

where z_1 denotes the number of teeth of the driving gear. Eq. (2) providing a physically grounded reference for spectral analysis. Frequency bands centered on mesh frequencies were constructed using an adaptive bandwidth to ensure robustness against minor speed variations. Spectral features were computed using Welch's method to estimate the power spectral density. Within the mesh band, spectral entropy was used to quantify the distribution of energy around the carrier frequency, while sideband-to-carrier ratios captured amplitude modulation effects associated with periodic tooth defects. A broadband high-frequency energy ratio was included to characterize impact-driven excitation.

Time-domain impulsiveness was quantified using Peak Dominance (PD),

$$PD = \frac{\max|s[n]|}{\sqrt{\frac{1}{L} \sum s[n]^2}} \quad (3)$$

Here, $s[n]$ represents the discrete vibration signal in a given time window. Eq. (3) measures transient amplitude excursions relative to signal energy. Local nonstationarity was

characterized using the coefficient of variation of short-time RMS values, reflecting intermittent bursts of energy typical of fault progression.

To enhance sensitivity to amplitude modulation, the signal was band-pass filtered around the primary mesh frequency and demodulated using the Hilbert transform. The envelope RMS,

$$RMS_e = \sqrt{\frac{1}{L} \sum e[n]^2} \quad (4)$$

quantified demodulated energy, while modulation depth at the rotational frequency was computed from the envelope spectrum to assess periodic amplitude fluctuations. In this expression, $e[n]$ denotes the analytic signal envelope obtained via Hilbert transform.

Impulsive behaviour was further characterized by detecting peaks in the envelope signal using a scale-adaptive prominence threshold. The impulse concentration relative to expected mesh interactions and the impulse energy normalized per rotational cycle,

$$E_{imp} = \frac{\sum e[n_k]^2}{f_{rot}} \quad (5)$$

where $e[n_k]$ are envelope samples at detected impulse locations. Eq. (5) were used to quantify impact density and severity. Periodicity consistency was evaluated through inter-peak interval variability.

To account for structural anisotropy and direction-dependent vibration response, cross-axis aggregation descriptors were introduced. Axis dominance captured directional concentration of vibration energy, while axis spread quantified inter-axis variability. Ratios between envelope energy, modulation depth and impulse energy were included to distinguish smooth rotational modulation from impact-dominated behaviour.

In this study, the faulty (NOK) state conditions correspond to experimentally observed defect scenarios in the gearbox system, associated with damage in specific gear components (e.g., defects in individual gears within the transmission stages). These faulty conditions are characterized by measurable deviations in vibration signal behaviour compared to healthy (OK) operation. In particular, defects manifest as increased energy and irregularities in frequency bands associated with gear-mesh frequencies and their harmonics, as well as changes in time-domain statistical descriptors such as impulsiveness and kurtosis. Therefore, the distinction between OK and NOK states is grounded in both the physical presence of gear defects and their observable signatures in the measured vibration signals.

Feature selection was performed separately for each rotation direction (RPM0-left and RPM1-right), acknowledging asymmetric dynamic behaviour under reversed torque transmission. For Logistic Regression models, selected features primarily emphasized mesh-band entropy, sideband modulation strength, impulsiveness (PD), envelope energy (RMS_e) and cross-axis variability. These features directly characterize amplitude modulation and localized impact effects, enabling linear discrimination between healthy (OK) and faulty (NOK) states.

In contrast, Isolation Forest models were trained exclusively on healthy data and relied more heavily on impulse concentration, impulse energy per rotation (E_{imp}), nonstationarity measures and axis-dominance descriptors. These features are particularly effective for describing deviations from nominal operational patterns without requiring explicit defect

labelling. Feature extraction was conducted at the window level for training dataset, while test files were processed on a case-by-case basis to ensure unbiased evaluation. All defect types were grouped into a single NOK class, resulting in a binary condition-monitoring framework consistent with industrial practice.

2.2. COMPUTATIONAL METHODS

The computational framework adopted in this study is schematically illustrated in Fig. 3. Following the extraction of physically interpretable vibration features from the tri-axial vibration signals (Section 2.1), machine-learning models were employed to classify gearbox operating conditions into healthy (OK) and faulty (NOK) states. Since gearbox dynamics differ depending on rotational direction, separate models were developed for left rotation (RPM0) and right rotation (RPM1). This direction-specific modelling ensures that characteristic vibration patterns associated with each operating condition are learned independently. For model development, a total of 40 gearbox experiments were utilized, comprising 25 healthy (OK) and 15 faulty (NOK) cases. These experiments were used for feature extraction and model fitting. An additional six independent experiments (three OK and three NOK) were reserved exclusively for final evaluation in order to assess generalization capability under previously unseen operating conditions.

The proposed framework combines two complementary modelling paradigms: an unsupervised anomaly detection approach based on Isolation Forest and a supervised probabilistic classifier based on Logistic Regression. Both models operate on the same engineered feature space but rely on fundamentally different learning principles. Isolation Forest models the statistical structure of healthy data and identifies deviations from this distribution, whereas Logistic Regression directly learns a discriminative boundary between OK and NOK samples.

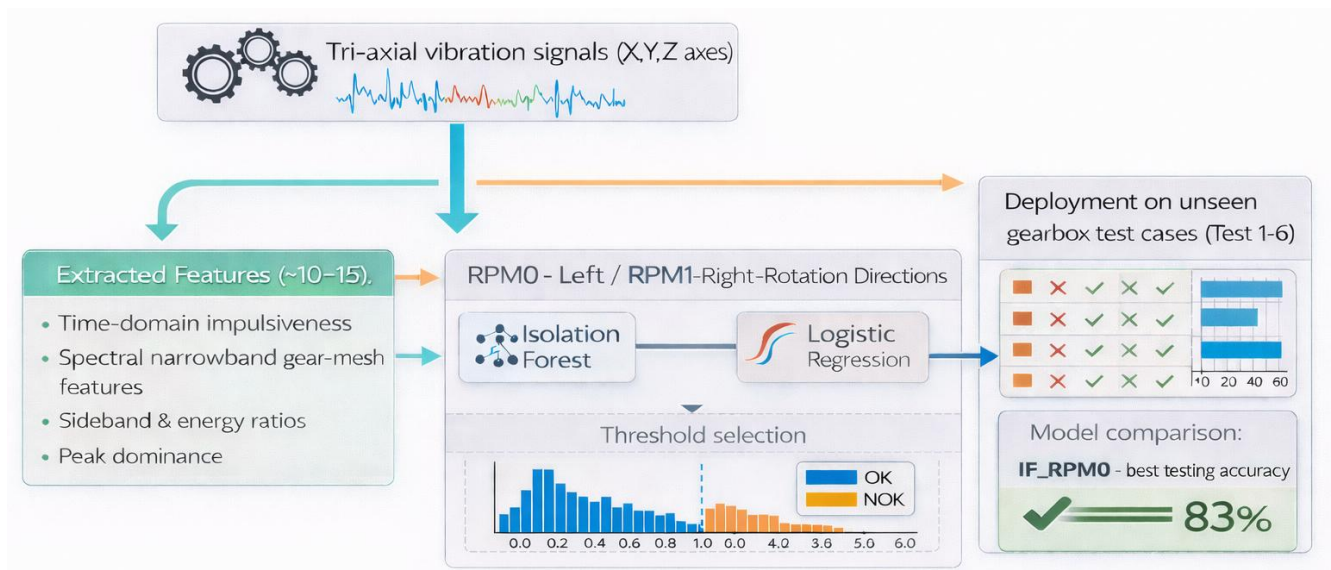


Fig. 3. Schematic overview of the proposed machine-learning framework for gearbox condition monitoring, showing the processing pipeline from tri-axial vibration signals (X, Y, Z) to feature extraction, direction-specific model training (RPM0 and RPM1), and final fault classification

2.2.1. MACHINE-LEARNING FRAMEWORK AND DATA PROCESSING

For each rotational direction, feature datasets were organized into a training dataset used for model fitting and an independent test set reserved exclusively for performance evaluation. The test data were not used during model development. Hyperparameter sensitivity analysis indicated that moderate variations in model parameters did not lead to substantial changes in classification performance. Therefore, standard model configurations were retained and decision thresholds were defined based on the observed anomaly score and probability distributions.

All extracted features are numerical descriptors derived from time-domain impulsiveness measures and frequency-domain gear-mesh characteristics. Prior to modelling, missing or undefined values were replaced with zero to ensure numerical stability. For the Logistic Regression model, features were standardized using z-score normalization defined as

$$x' = \frac{x - \mu}{\sigma} \quad (6)$$

where μ and σ denote the mean and standard deviation computed from the training data. Isolation Forest was applied directly to the raw feature vectors without scaling, as tree-based methods are insensitive to feature magnitude differences. Both models produce continuous outputs that require thresholding to obtain binary decisions. The Isolation Forest generates an anomaly score based on the expected path length required to isolate a sample within randomly constructed decision trees. These scores were robustly normalized using percentile-based scaling to the interval [0,1] and a fixed threshold of 0.55 was applied to distinguish anomalous behaviour. Logistic Regression produces the posterior probability of the faulty class defined as

$$P(NOK | x) = \frac{1}{1 + e^{-(w^T x + b)}} \quad (7)$$

and samples were classified as NOK when this probability exceeded 0.70. The decision thresholds used in this study were defined based on the empirical distributions of model outputs rather than through formal optimization procedures. Specifically, the threshold for the Isolation Forest model was selected using the normalized anomaly score distribution, while the Logistic Regression threshold was determined from the predicted probability distribution. In both cases, the thresholds were chosen to separate typical healthy behaviour from anomalous or defective patterns, as observed in the data. Importantly, no iterative tuning on a separate validation dataset was performed, in order to reduce the risk of overfitting and to maintain robustness in practical industrial applications. The selected thresholds therefore reflect a compromise between sensitivity and reliability, ensuring stable performance under varying operating conditions.

2.2.1. MODEL FORMULATION

Isolation Forest is an ensemble-based anomaly detection algorithm that isolates observations through recursive random partitioning of the feature space. Each tree randomly

selects a feature and a split value within its range, progressively partitioning the data until individual samples are isolated. The anomaly score is derived from the expected path length $E(h(x))$ required to isolate a sample x ; observations that are isolated with fewer splits are considered more anomalous. In this study, the model was trained exclusively on healthy (OK) samples in order to capture the statistical structure of normal gearbox operation. Separate ensembles were constructed for RPM0 and RPM1 conditions.

Logistic Regression, in contrast, is a supervised linear classifier that estimates the probability of the faulty class as a logistic transformation of a linear combination of the input features. Model parameters are obtained by minimizing a regularized logistic loss function with L2 regularization. Class-weight balancing was incorporated during training to account for potential imbalance between OK and NOK samples. Independent Logistic Regression models were trained for each rotational direction to accommodate direction-dependent vibration characteristics. Together, these two models provide complementary perspectives on gearbox health monitoring: Isolation Forest identifies deviations from learned normal behaviour, whereas Logistic Regression constructs an explicit discriminative boundary between healthy and faulty states. For clarity, the trained models are denoted as IF_RPM0 and IF_RPM1 for the Isolation Forest models corresponding to left and right rotation, respectively, and LR_RPM0 and LR_RPM1 for the corresponding Logistic Regression models. The outputs of both models are evaluated independently in the Summary section.

3. SUMMARY

3.1. MODEL OUTPUT CHARACTERISTICS

The effectiveness of the proposed machine-learning framework for gearbox condition monitoring was evaluated in two stages. First, model behaviour was analysed on the dataset used for model fitting to examine class separability and threshold behaviour. Second, the trained models were deployed on six unseen gearbox test cases to assess robustness and practical applicability under independent operating conditions.

Figure 4 illustrates the distributions of model outputs for both rotation directions. Panels (a) and (b) present the normalized Isolation Forest anomaly scores for RPM0 and RPM1, while panels (c) and (d) show the Logistic Regression predicted probabilities $P(\text{NOK})$. Blue and orange histograms correspond to OK and NOK samples, respectively, and dashed vertical lines indicate predefined classification thresholds. For both directions, the Logistic Regression model provides clearer separation between healthy and faulty samples. NOK samples are predominantly concentrated at higher predicted probabilities, while OK samples remain near zero. The fixed probability threshold of 0.70 ensures conservative classification with reduced false-positive decisions. In contrast, the Isolation Forest anomaly scores exhibit greater overlap between classes. Although NOK samples generally yield higher anomaly scores, the separation remains less distinct, particularly for RPM1. The fixed anomaly-score threshold of 0.55 provides a compromise between sensitivity to abnormal behaviour and stability in recognizing normal operating conditions.

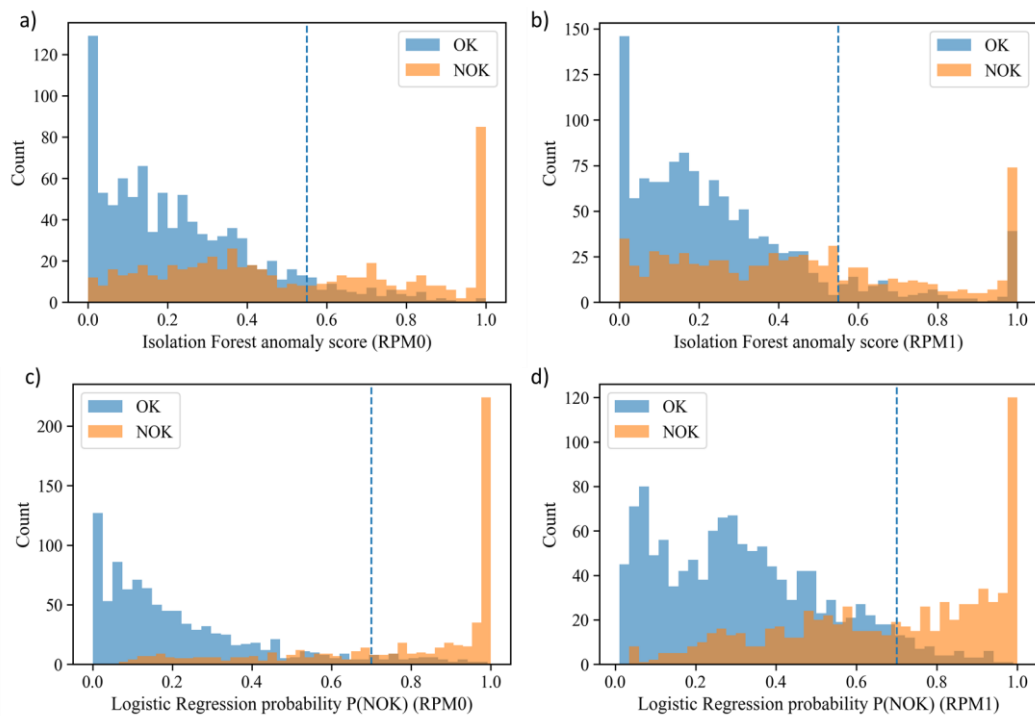


Fig. 4. Distribution of model outputs for both rotation directions. (a–b) Isolation Forest anomaly scores for RPM0 and RPM1. (c–d) Logistic Regression predicted probabilities (NOK). Blue and orange histograms represent OK and NOK samples; dashed lines indicate classification thresholds

3.2. CLASSIFICATION PERFORMANCE ON THE TRAINING DATASET

Figure 5 presents the normalized confusion matrices together with the aggregated performance metrics for both rotation directions. In addition to the class-wise prediction structure, the comparison includes balanced accuracy, F1-score for the NOK class and recall for the NOK class, enabling an assessment that accounts for class imbalance and fault-detection sensitivity.

For the RPM0 direction, Logistic Regression achieves the highest balanced accuracy and F1-score, indicating a strong capability to separate healthy and faulty operating conditions when labelled data are available for supervised training. The elevated F1-score reflects a favourable balance between precision and recall for the NOK class, demonstrating that the model effectively captures discriminative patterns associated with gearbox faults. In contrast, while Isolation Forest correctly identifies a large proportion of OK samples, its recall for the NOK class is lower, particularly in the RPM1 case. This behaviour is consistent with its unsupervised nature, as the model is trained exclusively on normal-operation data and therefore identifies faults indirectly through deviation from learned normal patterns.

Across both rotation directions, Logistic Regression consistently provides higher sensitivity to faulty conditions, as evidenced by improved NOK recall values. Isolation Forest, however, exhibits greater stability in recognizing normal operation, reflected in its high true-negative rates. This complementary behaviour highlights the distinction between supervised probabilistic classification and anomaly-based detection approaches. It should be emphasized that all model parameters and classification thresholds were predefined and not subjected to

iterative optimization or hyperparameter tuning. Consequently, the reported metrics represent direct baseline performance under fixed decision criteria, thereby providing an unbiased comparison of the two modelling strategies.

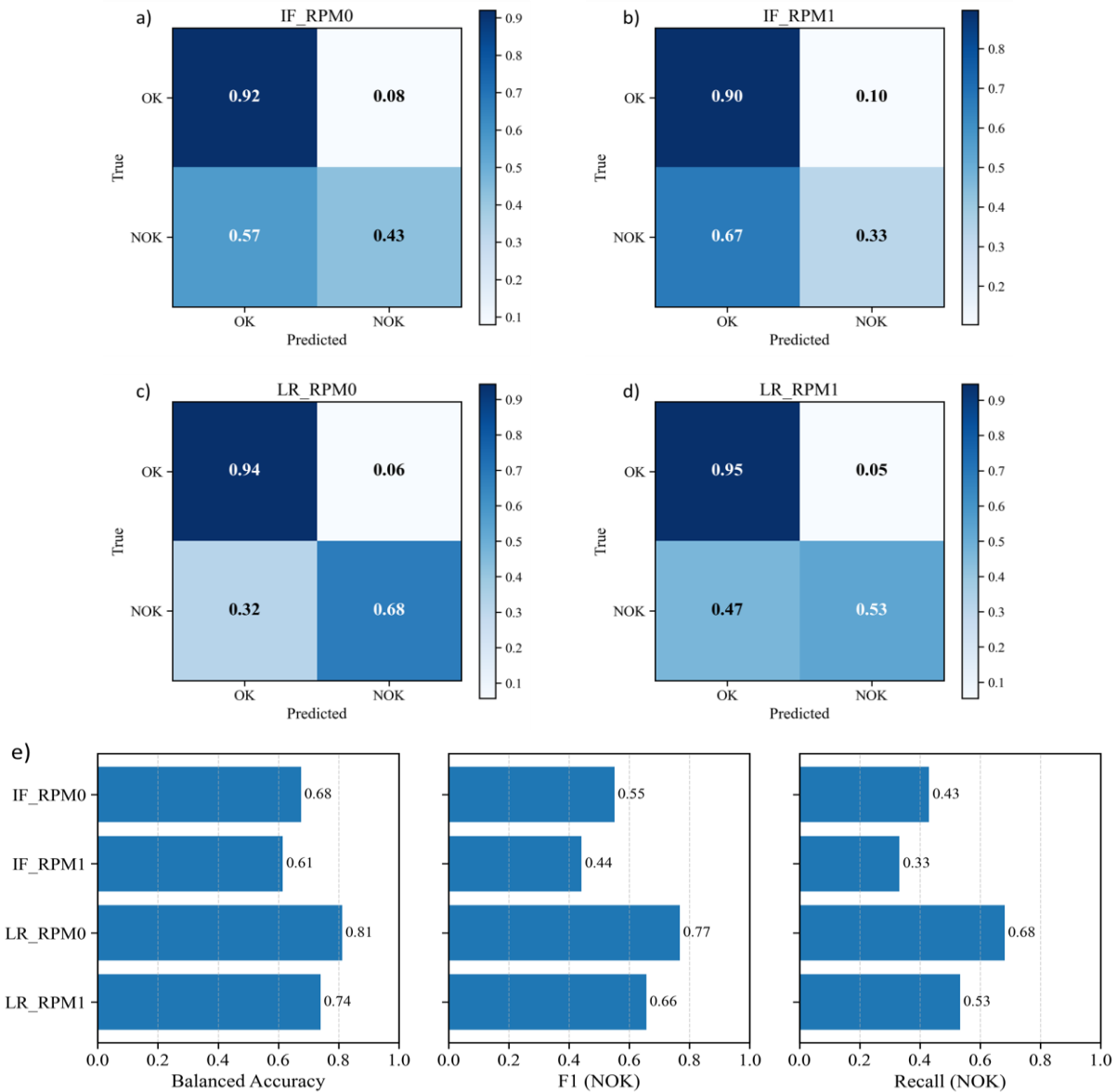


Fig. 5. Normalized confusion matrices and performance metrics for the training dataset. (a–b) Isolation Forest results for RPM0 and RPM1. (c–d) Logistic Regression results for RPM0 and RPM1. (e) Summary of balanced accuracy, F1-score (NOK class), and recall (NOK class) for all model–direction combinations

3.3. DEPLOYMENT ON UNSEEN GEARBOX TEST CASES

The generalization capability of the trained models was assessed on six independent gearbox test cases. Figure 6a summarizes prediction correctness for each model–direction combination, while Fig. 6b compares OK accuracy, NOK accuracy and overall accuracy.

The independent test set consisted of three healthy (OK) and three faulty (NOK) gearbox cases, resulting in a balanced evaluation protocol.

The IF_RPM0 model achieved the highest overall testing accuracy (83%), correctly identifying all healthy cases and detecting the majority of faulty cases. The performance of RPM1-based models was comparatively lower, indicating that vibration characteristics under right-rotation conditions may be more challenging for separation in the selected feature space.

Although Logistic Regression demonstrated strong performance on the fitting dataset, its testing results show higher variability compared to the Isolation Forest in RPM0. This suggests that supervised classification may be more sensitive to distribution shifts between training and deployment conditions.



Fig.6. Deployment results on six unseen gearbox test cases. a) Prediction correctness matrix for Isolation Forest (IF) and Logistic Regression (LR) models for RPM0 and RPM1. b) Comparison of OK-class accuracy, NOK-class accuracy, and overall accuracy across models

3.4. ENGINEERING IMPLICATIONS AND LIMITATIONS

From an engineering perspective, the proposed diagnostic framework offers several practical advantages for industrial gearbox monitoring. The feature set is compact and physically interpretable, allowing a direct connection between the observed vibration characteristics and the mechanical behaviour of the gearbox. Such interpretability is particularly important in industrial environments where maintenance decisions must be supported by understandable diagnostic indicators rather than purely black-box predictions.

Another important advantage of the proposed approach is its computational efficiency. Both Isolation Forest and Logistic Regression are lightweight machine-learning models that require relatively modest computational resources compared to deep learning architectures. This makes the framework suitable for integration into industrial monitoring systems where rapid analysis and near real-time decision making are required. The use of direction-specific models represents an additional practical benefit. Gearbox dynamics may vary depending on the direction of torque transmission and mechanical loading conditions. By training independent models for RPM0 and RPM1, the framework is able to capture direction-dependent vibration behaviour more effectively, improving the reliability of the diagnostic process. Despite these advantages, several limitations should be acknowledged. The experimental dataset used in this study consists of a relatively limited number of gearbox experiments. Although independent test cases were used to assess generalization capability, larger datasets covering broader operational conditions would further improve statistical robustness. In addition, the present study focuses on binary classification between healthy (OK) and faulty (NOK) states. While this approach reflects common industrial practice for quality control, future work could extend the framework toward multi-class fault identification in order to distinguish between specific defect types.

Another limitation concerns the range of operating conditions considered in the experiments. The measurements were conducted at a fixed rotational speed and under controlled testing conditions. In real industrial environments, gearboxes may operate under varying speeds, loads, and environmental influences, which may affect vibration characteristics. Future studies should therefore investigate the robustness of the proposed framework under a wider range of operating regimes. Overall, the results demonstrate that combining physically motivated vibration features with simple and interpretable machine-learning models provides a promising approach for practical gearbox condition monitoring. Further work will focus on extending the methodology to larger datasets, incorporating additional operating conditions, and exploring its integration into automated predictive maintenance systems.

Overall, the results demonstrate that combining physically motivated vibration features with simple and interpretable machine-learning models provides a promising approach for practical gearbox condition monitoring. Several strategies may be considered to further improve the predictive performance of the proposed models. First, increasing the size and diversity of the dataset, particularly by including a broader range of operating conditions such as variable rotational speeds, load levels, and environmental influences, would enhance model generalization. Second, refinement of the feature engineering process, including the incorporation of additional gear-mesh-related descriptors and more detailed time-frequency representations, may improve sensitivity to subtle fault signatures. Third, the integration of more advanced machine-learning approaches, such as ensemble methods or nonlinear models, could provide improved discrimination capability while maintaining robustness. Finally, combining the outputs of anomaly detection and supervised classification within a unified decision framework represents a promising direction for improving reliability and reducing misclassification in practical industrial applications. These improvements will be considered in future work, with particular focus on enhancing robustness under realistic industrial operating conditions and extending the applicability of the proposed framework.

It should be noted that the vibration data used in this study were obtained from industrial factory tests and are not publicly available due to confidentiality constraints. However, several open-access datasets exist that provide comparable vibration signal measurements for gearbox condition monitoring and fault diagnosis. In particular, recently published datasets under variable working conditions [18] offer representative data that can be used for training and validating similar machine-learning approaches. The methodology proposed in this work is directly transferable to such datasets, supporting reproducibility and facilitating further development of data-driven diagnostic frameworks.

4. CONCLUSIONS

A hybrid machine-learning framework for car gearbox quality control and fault detection was proposed and validated using tri-axial vibration features. The methodology combines anomaly detection through Isolation Forest with supervised classification via Logistic Regression, implemented separately for two rotational directions. The results demonstrate that the proposed approach is capable of effectively distinguishing between healthy (OK) and faulty (NOK) operating conditions. Logistic Regression showed strong class discrimination on the fitting dataset, while the Isolation Forest trained on RPM0 provided the most robust generalization on unseen test cases, achieving the highest overall deployment accuracy.

A key finding of this work is that anomaly-based modelling of healthy operation enables reliable fault detection, particularly in situations where labelled faulty data are limited. In addition, physically interpretable, gear-mesh-related features were shown to provide sufficient discriminatory information when combined with lightweight machine-learning models. The use of direction-specific models (RPM0 and RPM1) further improved diagnostic consistency by capturing differences in vibration behavior associated with the direction of operation. Moreover, the hybrid decision strategy enhances robustness by combining anomaly detection and probabilistic classification, resulting in a stable and interpretable diagnostic outcome suitable for industrial applications.

Overall, the results highlight that combining domain-informed feature engineering with simple and computationally efficient machine-learning techniques provides a reliable and scalable solution for gearbox condition monitoring. Future work will focus on extending the framework to more complex operating conditions, larger datasets, and multi-class fault identification.

ACKNOWLEDGEMENTS

We acknowledge support from the European Union Horizon 2020 research and innovation program under grant agreement no. 857470 and from the European Regional Development Fund via the Foundation for Polish Science International Research Agenda PLUS program grant No. MAB PLUS/2018/8. The authors also gratefully acknowledge ABM Greiffenberger Polska Sp. z o.o. for providing access to industrial gearbox systems and for conducting the factory tests that enabled the experimental validation presented in this study.

REFERENCES

- [1] AKMEASE R., EASTHAM J., 1992, *Design of Permanent Magnet at Linear Motors*, IEEE Trans. Magn., 28/5, 302–306.
- [2] YAQOB M.F., GOLLEE C., SCHENKE C., IHLENFELDT S., 2025, *Concept of Developing a Gear Selection Tool for Improved Accuracy in Industrial Robotics*, Journal of Machine Engineering. 25/2, 138, <https://doi.org/10.36897/jme/204359>.
- [3] DINH V., TRAN H., VU D., VU D., VU N., LUU T.A., 2024, *Using TOPSIS Method for Solving Multi-Objective Optimization of a Two-Stage Helical Gearbox with First Stage Double Gear-Sets*, Journal of Machine Engineering, 24/4, 41. <https://doi.org/10.36897/jme/196771>.
- [4] STEMBAŁSKI M., JANKOWSKI T., ROSZKOWSKI A., SKOCZYNSKI W., 2025, *Gear Wear Mechanisms, Monitoring Techniques and Their Potential Use in Gear Predictive Maintenance*, Journal of Machine Engineering, 25/2, 43, <https://doi.org/10.36897/jme/206045>.
- [5] HORVATH K., ZELEI A., 2024, *Gearbox Fault Diagnosis Using Industrial Machine Learning Techniques*, Engineering Proceedings 79/1, 36, <https://doi.org/10.3390/engproc2024079036>.
- [6] TUYET-DOAN V.-N., CHOI M., PARK G. 2025. *XGBoost Method-Based Gearbox Fault Diagnosis Using Time-Domain Signal Under Road Vehicle Characteristics*, Electronics, 14(23), 4736. <https://doi.org/10.3390/electronics14234736>.
- [7] MAMMADLI B., YAZICI C., GÜRBÜZ M., KOCAMAN R., JAVIER DOMINGUEZ-GUTIERREZ F., ÖZKAN F.M., 2025, *A Data-Driven Machine Learning Approach for Predicting Axial Load Capacity in Steel Storage Rack Columns*, Results in Engineering 28, 107475.
- [8] RZECZKOWSKI J., 2026, *Application of an Unsupervised Deep Learning Framework for Acoustic Emission-Based Characterization of Delamination Process in Composite Laminates*, Composites Science and Technology 276, 111507.
- [9] KARPIŃSKI P., RZECZKOWSKI J., 2026, *Comparison of Unsupervised Machine Learning Segmentation Algorithms in the Analysis of Unmanned Aerial Vehicle - Based Multispectral Crop Images*, Advances in Science and Technology Research Journal, 20/2, 94–113, <https://doi.org/10.12913/22998624/210743>.
- [10] SURYAWANSHI S., GAVALI S., DARWAI V., PATIL S., WANKHEDE A., BORSE A., 2024, *Gearbox Fault Diagnosis: A Comparative Study of Machine Learning and Deep Learning Approaches*, Ranganathan, G., Papakostas, G.A., Shi, Y. (eds) Inventive Communication and Computational Technologies, ICICCT 2024, Lecture Notes in Networks and Systems, 23, Springer, Singapore. https://doi.org/10.1007/978-981-97-7710-5_25.
- [11] HASSAN M., SADEK A., ATTIA M.H., 2022, *Intelligent Cyber-Physical Monitoring and Control of I4.0 Machining Systems - an Overview and Future Perspectives*, Journal of Machine Engineering. 22/1, 5–24. <https://doi.org/10.36897/jme/147157>.
- [12] ERNESTO P., FERNANDEZ D., RODRIGUEZ-PRIETO A., 2026, *Diagnostic Analytics Powered by IOT and Machine Learning for the Fault Evaluation of a Heavy-Industry Gearbox*, Machines, 14/2, 187, <https://doi.org/10.3390/machines14020187>.
- [13] NGUYEN V-T., DIEP Q.B., 2025, *Vibration-Based Gearbox Fault Diagnosis Using a Multi-Scale Convolutional Neural Network with Depth-Wise Feature Concatenation*, PLoS One 20/7, e0324905, <https://doi.org/10.1371/journal.pone.0324905>.
- [14] SURYAWANSHI S., GHORPADE U., BHOSALE D.G., BHOSALE A.P., PATIL M.A., 2025, *Application of Machine Learning and Deep Learning in Gearbox Fault Diagnosis: A Comprehensive Review*, Gunjan, V.K., Zurada, J.M. (eds) Proceedings of 5th International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications. ICMISC 2024, Lecture Notes in Networks and Systems, 1182. Springer, Singapore, https://doi.org/10.1007/978-981-97-8865-1_29.
- [15] CHUAN LI, RENE-VINICIO SANCHEZ, GROVER ZURITA, MARIELA CERRADA, DIEGO CABRERA, RAFAEL E. VASQUEZ, 2016. *Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signals*. Mechanical Systems and Signal Processing 76, 283.
- [16] ZHANG C., GUO Z., LI C., 2025, *Unsupervised Anomaly Detection for Gearboxes Based on the Deep Convolutional Support Generative Adversarial Network*, Sci. Rep. 15, 20977, <https://doi.org/10.1038/s41598-025-08835-0>.
- [17] TANG H., WANG H., LI C., 2025, *Explainable Fault Diagnosis Method Based on Statistical Features for Gearboxes*, Engineering Applications of Artificial Intelligence, 148, 110503.
- [18] CHEN S., LIU Z., HE X., ZOU D., ZHOU D., 2024, *Multi-Mode Fault Diagnosis Datasets of Gearbox Under Variable Working Conditions*, Data in Brief 54, 110453