Wujud: Journal of Engineering and Technology
2026, 1 (1), 93-108

Deep Learning Based Fault Diagnosis for Predictive Maintenance of Industrial Rotating Machinery Using Vibration Signals

Aditya Bagus Santoso[1]*; Rina Oktavia Sari[2]; Rafi Alfarizi Hidayat[3]; Maya Fitri Handayani[2]
[1]Department of Mechanical Engineering, Faculty of Engineering, Universitas Indonesia, Depok, Indonesia
[2]Department of Electrical Engineering, Faculty of Engineering, Universitas Indonesia, Depok, Indonesia
[3]Department of Industrial Engineering, Faculty of Engineering, Universitas Indonesia, Depok, Indonesia

Article History
Received: 2026-01-28
Revised: 2026-03-03
Accepted: 2026-04-22
Published: 2026-05-31

Keywords:
Predictive maintenance; Rotating machinery; Vibration signal; Bearing fault diagnosis; Deep learning.

DOI:
10.xxxx/wujud.volx.issx.artx

Corresponding author:
Aditya Bagus Santoso
adityabagussantoso@gmail.com

Other author email(s):
rinaoktaviasari@gmail.com
rafialfarizihidayat@gmail.com
mayafitrihandayani@gmail.com

Paper type:
Research article

Published by PT Penerbit Velora Indonesia

Abstract

Purpose – This study evaluates vibration-based bearing fault diagnosis using a Deep Multilayer Perceptron (Deep MLP) and compares its performance with SVM-RBF and Random Forest using identical time-domain vibration features.

Methodology – A quantitative benchmark design used 128 balanced CWRU-derived observations representing Normal, Ball Fault, Inner Race Fault, and Outer Race Fault. Nine time-domain features were analyzed. Performance was estimated through repeated stratified 5-fold cross-validation with 10 repetitions, accuracy, macro precision, macro recall, macro F1-score, and Wilcoxon signed-rank tests.

Findings – Deep MLP achieved mean accuracy and macro F1-score of 99.53%. SVM-RBF and Random Forest each achieved 100.00%. The difference was statistically significant at $p = 0.0256$; therefore, vibration-feature discriminability was supported, while Deep MLP superiority was not supported.

Implications – Highly discriminative engineered vibration features can allow conventional classifiers to match or exceed deeper neural models with lower computational complexity.

Originality – The study provides a controlled comparison of Deep MLP, SVM-RBF, and Random Forest using identical features and repeated validation partitions, showing that deep-learning superiority depends on input representation and task structure.

Cite this article:
Santoso, A. B., Sari, R. O., Hidayat, R. A., & Handayani, M. F. (2026). Deep learning based fault diagnosis for predictive maintenance of industrial rotating machinery using vibration signals. Wujud: Journal of Engineering and Technology, x(x), xx–xx. https://doi.org/10.xxxx/wujud.volx.issx.artx

Introduction

Industrial rotating machinery is essential to manufacturing, energy production, transportation, petrochemical processing, and other continuous industrial operations. Motors, bearings, gearboxes, pumps, compressors, turbines, and shaft systems operate under repeated mechanical loads and are therefore exposed to gradual deterioration. Unexpected component failure can interrupt production, increase maintenance expenditure, reduce equipment availability, and create operational risks. Das et al. (2023) explain that data-driven fault analysis has become increasingly important because complex machinery faults cannot always be detected reliably through manual inspection or predefined diagnostic rules. Vibration monitoring is particularly useful because changes in mechanical condition produce measurable changes in dynamic responses. Bagri et al. (2024) identified vibration analysis as one of the central approaches for intelligent diagnosis and prognosis of rotating machinery. Deep learning has further expanded this capability by allowing complex relationships between monitoring signals and equipment conditions to be learned

P ISSN xxxx-xxxx | E ISSN xxxx-xxxx

automatically. Ali et al. (2024) showed that deep learning-based condition monitoring can provide a technical basis for predictive maintenance of rotational machinery.

The need for reliable fault diagnosis is increasingly relevant as industrial systems adopt automated and data-driven maintenance practices. Conventional corrective maintenance responds after failure has occurred, while scheduled preventive maintenance may replace components that remain functional. Predictive maintenance instead uses condition information to identify deterioration before failure becomes critical. Atmaji et al. (2025) demonstrated that changes in shaft misalignment can be identified from vibration characteristics using machine-learning and deep-learning techniques. Industrial applications also involve a substantial imbalance between normal operating observations and actual fault records. Anshori and Dhini (2025) addressed this issue in rotating machinery within a petrochemical setting and demonstrated the potential of artificial intelligence for detecting abnormal equipment states. Vibration information collected from different machine positions can also support fault identification under varying dynamic conditions. Almutairi et al. (2024) found that combining vibration information can reduce redundant parameters while maintaining fault-sensitive characteristics. These developments indicate that automated vibration diagnosis can strengthen condition-based maintenance decisions.

Vibration signals contain physical information related to mechanical interaction within rotating components. Localized bearing defects, imbalance, misalignment, and gear damage can change amplitude, impulsiveness, periodicity, spectral energy, and statistical distributions. Li and Bai (2024) emphasize that vibration-based bearing diagnosis generally includes signal preprocessing, fault-feature extraction, and condition identification. Time-domain characteristics such as root mean square, standard deviation, kurtosis, skewness, crest factor, and peak amplitude provide compact numerical representations of dynamic behavior. Jiang et al. (2024a) showed that signal-processing features can support machine-learning models for mechanical fault diagnosis and prediction. Time-frequency methods provide another representation when signal characteristics are non-stationary. Zhang and Deng (2023) demonstrated that Short-Time Fourier Transform can expose bearing-fault patterns that can subsequently be learned by convolutional networks. Nevertheless, time-domain features remain attractive when the objective is computationally efficient diagnosis using compact and interpretable variables.

Deep learning reduces the dependence on fixed diagnostic thresholds because nonlinear relationships among vibration characteristics can be learned directly from training data. Su et al. (2024a) identified convolutional neural networks, recurrent networks, autoencoders, transfer learning, and Transformer architectures as major developments in rotating-machinery diagnosis. Hybrid architectures can combine complementary feature-learning mechanisms. Song et al. (2024) integrated convolutional and bidirectional temporal learning to improve bearing diagnosis under multiple working conditions and limited training samples. Attention-based methods further extend this capability by assigning different weights to diagnostically important information. Zhang et al. (2024b) demonstrated that Transformer-based processing can capture complex dependencies in vibration time series. However, increasingly sophisticated models do not automatically provide superior performance when the input variables already contain strongly discriminative fault information.

This issue is important because high benchmark accuracy can conceal limitations in generalization and model selection. Matania et al. (2024) found a persistent gap between high-performing laboratory fault-diagnosis models and the requirements of critical rotating machinery in actual industrial environments. Class imbalance can also distort performance interpretation when models achieve high overall accuracy but perform poorly on less frequent fault categories. Prawin (2025) showed that bearing-diagnosis performance changes substantially under imbalanced-data conditions and that macro-level evaluation is important for assessing stability. Operational variation introduces another challenge because signals recorded at different speeds, loads, or machines may follow different distributions. Rezazadeh et al. (2025) demonstrated that domain adaptation can improve rotor fault diagnosis across operational variability. These findings suggest that model evaluation should consider not only classification accuracy but also class-balanced performance and validation stability.

Another unresolved question concerns whether complex deep architectures are always necessary for vibration-based fault diagnosis. Kibrete et al. (2025) reported strong performance from a hybrid CNN-LSTM architecture, illustrating the effectiveness of automated spatial and temporal feature learning. However, conventional machine-learning algorithms can remain competitive when fault-sensitive features have already been extracted. Nguyen et al. (2025) showed that performance differences among machine-learning and deep-learning methods depend strongly on the selected input representation and diagnostic task. Time-domain feature analysis also provides a computationally efficient basis for neural classification. Jorge et al. (2024) demonstrated that an MLP can identify multiple rotating-machine faults using features such as RMS, variance, standard deviation, peak-to-peak amplitude, and shape factor. The empirical question is therefore not simply whether deep learning works, but whether a deeper nonlinear classifier provides a measurable advantage over established machine-learning alternatives when all models receive the same vibration features.

Based on this gap, the present study evaluates fault diagnosis using a Deep Multilayer Perceptron applied to time-domain vibration features derived from the Case Western Reserve University bearing dataset. The model is compared directly with Support Vector Machine using a radial basis function kernel and Random Forest under identical validation partitions. The analysis focuses on four bearing conditions: Normal, Ball Fault, Inner Race Fault, and Outer Race Fault. Classification performance is evaluated using accuracy, macro precision, macro recall, macro F1-score, confidence intervals, and repeated cross-validation. This design addresses two research questions: RQ1: To what extent can time-domain vibration features discriminate normal and faulty bearing conditions using a Deep MLP model? RQ2: Does Deep MLP provide significantly better diagnostic performance than SVM-RBF and Random Forest under an identical feature and validation framework? The resulting evidence positions fault classification as an empirical diagnostic layer that can support broader predictive-maintenance systems rather than treating classification accuracy alone as a complete maintenance solution.

Literature Review

Deep Learning for Rotating Machinery Fault Diagnosis

The development of intelligent fault diagnosis reflects a transition from manually defined diagnostic rules toward models that learn nonlinear patterns from condition-monitoring data. Deep learning is particularly useful when interactions among features cannot be represented adequately by simple linear boundaries. Wang et al. (2023) demonstrated that bearing fault information can be modeled through graph formulation and graph convolution, illustrating that diagnostic relationships extend beyond individual vibration measurements. Graph-based methods can capture relationships among observations or features and therefore provide an alternative to conventional Euclidean neural structures.

Attention mechanisms have further improved the ability of diagnostic networks to emphasize important information. Ma et al. (2024) developed a self-attention Legendre graph convolutional network for rotating-machinery fault diagnosis and reported improved adaptability across operating conditions. Multi-scale representations provide another mechanism for capturing fault information that appears at different structural resolutions. Wei et al. (2024) introduced a multi-scale graph Transformer and demonstrated that combining graph structure with self-attention can improve rolling-bearing diagnosis. These approaches show that modern diagnostic research increasingly focuses on the quality of representation rather than only increasing network depth.

Limited training data remain a major challenge for complex deep networks. Jiang et al. (2024b) developed a dynamic-edge graph convolution approach to improve diagnosis under few-shot conditions. Their findings indicate that adaptive feature relationships can reduce some of the limitations created by small fault datasets. Cross-domain variability creates another challenge because fault patterns can change as severity or operating condition changes. Gao et al. (2025) proposed a Transformer-based framework for rotating machinery with varying damage degrees and demonstrated the importance of multidomain feature integration.

Hybrid CNN and attention structures are also increasingly used for bearing diagnosis. Hu (2025) combined CNN-based local feature extraction with Transformer-based global learning to improve classification under different public datasets. Temporal dependency remains relevant because vibration characteristics evolve sequentially rather than appearing as isolated measurements. Yang et al. (2025) integrated CNN, BiLSTM, and multi-head self-attention for rotor-motor bearing diagnosis and reported strong classification performance under variable speeds. These studies establish the effectiveness of advanced deep architectures but also highlight their increased computational and methodological complexity.

Vibration Features for Bearing Condition Monitoring

The quality of fault classification depends on the information represented by the input variables. Single-sensor vibration observations may contain incomplete information, especially when the propagation path between a fault and a sensor changes. Xie et al. (2023) showed that multi-sensor spatiotemporal representations can improve gearbox fault diagnosis by combining complementary vibration information. Even when a single sensing channel is used, signal representation strongly affects classification because different features emphasize different aspects of mechanical behavior.

Statistical time-domain features offer a compact way to quantify vibration behavior. RMS reflects overall vibration energy, standard deviation describes variability, kurtosis is sensitive to impulsive characteristics, and crest factor relates peak magnitude to RMS level. Zhang et al. (2023) demonstrated that transforming vibration characteristics into structured representations can improve gearbox classification. Zhang et al. (2024a) similarly showed that visual and attention-enhanced representations can highlight diagnostically important bearing information. These findings do not eliminate the relevance of statistical features. Instead, they indicate that feature representation should match the complexity of the diagnostic task.

Signal decomposition is useful when mechanical responses contain overlapping or non-stationary components. Dong et al. (2024) combined empirical wavelet transformation with an attention-enhanced convolutional network to extract bearing-fault information. Multi-scale processing can also capture information that occurs at different temporal resolutions. Kang et al. (2023) demonstrated that multi-scale convolution improves feature extraction for rotating machinery operating under complex conditions. These methods are particularly relevant when raw signals must be processed automatically.

Noise can substantially reduce the visibility of fault-sensitive patterns. Li et al. (2025) developed a Transformer-based diagnostic approach designed for strong-noise conditions and showed that robust feature learning remains necessary when signal quality declines. Load variation can produce a similar distribution change because vibration amplitude and frequency composition depend on operating conditions. Ahsan et al. (2025) found that diagnostic accuracy differed across load levels even when the same hybrid architecture was used. Consequently, operating conditions should be considered when generalizing results from controlled vibration datasets.

The current study uses engineered time-domain features rather than raw vibration waveforms. This design provides a direct test of whether a Deep MLP can learn nonlinear class boundaries from a low-dimensional feature space. The approach also permits a fair comparison with SVM and Random Forest because all models receive exactly the same variables. This distinction is important because performance differences can otherwise reflect different preprocessing pipelines rather than differences in classifier capability.

Machine Learning, Deep Learning, and Predictive Maintenance

Fault classification represents one layer of predictive maintenance. A broader predictive-maintenance system can combine detection, diagnosis, degradation assessment, and remaining useful life estimation. Tran et al. (2024) demonstrated that multiple vibration representations can be integrated within a deep architecture for remaining useful life prediction. Such prognostic methods extend beyond classification because they estimate how long a component can continue operating before intervention is required.

Accurate diagnosis remains necessary before reliable prognostic decisions can be made. Mayo et al. (2024) showed that vibration-based deep learning can distinguish healthy and faulty gearbox conditions with strong accuracy. Limited fault data, however, can make severity classification more difficult. Mrabti et al. (2025) addressed this issue through a VMD-LSTM framework designed for fault detection and severity classification under limited-data conditions.

Explainability and efficient deployment are becoming increasingly important as diagnostic models move closer to industrial implementation. Jweri et al. (2025) combined signal transformation with an alternative neural architecture and emphasized the relevance of interpretable fault diagnosis for predictive maintenance. Hybrid decomposition and attention mechanisms provide another route for improving fault representation. Wang et al. (2025) combined variational mode decomposition, CNN, and Transformer components to improve rolling-bearing classification.

Industrial models must also respond to changing operating conditions. Su et al. (2024b) used multi-feature extraction and transfer learning to diagnose rotating machinery under variable operating conditions. Their work demonstrates that feature-based models and deep transfer mechanisms can complement one another when the source and target distributions differ. The implication is that neither engineered features nor deep representation learning should be treated as universally superior. Their value depends on data structure, operating variability, computational constraints, and the purpose of the maintenance system.

Research Gap and Study Positioning

Existing research provides strong evidence that deep learning can classify rotating-machinery faults accurately, but much of the literature focuses on increasingly complex CNN, recurrent, Transformer, or graph architectures. This emphasis can obscure an important practical question: whether complex feature learning remains necessary once fault-sensitive statistical variables have already been extracted. Siavash-Abkenari et al. (2024) showed that statistical analysis can be integrated effectively with deep learning for bearing condition monitoring. Their findings support a diagnostic framework in which statistical information and neural classification are complementary rather than mutually exclusive.

A second gap concerns fair model comparison. Different studies often compare algorithms using different preprocessing, input representations, or data partitions, making it difficult to determine whether performance differences arise from the classifier itself. Liu et al. (2024) showed that distribution shifts across conditions and machines complicate direct interpretation of diagnostic performance. A controlled comparison should therefore keep the input representation and validation partitions constant across competing models.

The present study addresses these gaps by providing Deep MLP, SVM-RBF, and Random Forest with the same nine time-domain vibration features and evaluating them across identical repeated stratified folds. The study does not assume that deep learning must outperform conventional methods. Instead, it empirically tests that expectation. This positioning is important because a simpler classifier may be preferable when it achieves equal or better performance using fewer computational resources. The resulting analysis contributes to predictive-maintenance research by identifying the circumstances under which a deep nonlinear classifier provides, or fails to provide, an advantage after feature engineering has already been performed.

Hypothesis Development

Time-domain vibration features summarize mechanical behavior into variables that can distinguish machine states. RMS captures changes in vibration energy, while standard deviation represents signal dispersion. Kurtosis is sensitive to impulsive events that can emerge from localized bearing damage. Crest factor provides additional information about the relationship between signal peaks and overall vibration magnitude. Jorge et al. (2024) demonstrated that these types of time-domain parameters can support MLP-based multi-fault classification in rotating systems. Almutairi et al. (2024) also showed that vibration-derived parameters retain diagnostic information across different machine conditions.

The selected CWRU-derived feature space contains nine statistical characteristics that represent amplitude, variability, impulsiveness, and waveform structure. A nonlinear neural classifier should therefore be able to learn boundaries between Normal, Ball Fault, Inner Race Fault, and Outer Race Fault observations. The first hypothesis is formulated as follows:

H1: Time-domain vibration features provide sufficient discriminative information for the Deep MLP model to classify normal and faulty bearing conditions with high diagnostic performance.

Deep MLP introduces multiple nonlinear transformations between input variables and output classes. This structure can theoretically represent complex relationships that conventional classifiers may not capture with equivalent flexibility. Siavash-Abkenari et al. (2024) showed that deep learning can complement statistical bearing-condition information. Su et al. (2024a) similarly observed that deep architectures have become important because of their capacity to learn complex diagnostic mappings.

However, conventional machine-learning algorithms remain strong competitors when input features are already highly discriminative. Nguyen et al. (2025) demonstrated that the relative performance of machine-learning and deep-learning models varies according to dataset and representation. Random Forest can form nonlinear ensemble decision boundaries without neural optimization, whereas an RBF-based SVM can separate complex classes through kernel mapping. The empirical comparison is therefore necessary rather than assuming that network depth automatically produces superior diagnosis.

Based on the theoretical flexibility of multilayer neural learning, the second hypothesis proposes an expected performance advantage for Deep MLP:

H2: The Deep MLP model achieves significantly higher fault-diagnosis performance than SVM-RBF and Random Forest when the models are evaluated using identical time-domain vibration features and validation partitions.

H1 and H2 address different questions. H1 evaluates the diagnostic adequacy of vibration features when processed by Deep MLP. H2 evaluates comparative model superiority. A high Deep MLP accuracy therefore does not automatically support H2. The second hypothesis is supported only if the neural model significantly exceeds the conventional baselines under the same evaluation procedure.

Research Methods

Research Design and Dataset

This study employed a quantitative experimental benchmark design. The analysis used a CWRU-derived bearing vibration feature dataset originating from the Case Western Reserve University Bearing Data Center. The final analytical dataset represented a 0.007-inch bearing fault configuration at a motor load of 1 hp. Four classes were included: Normal, Ball Fault, Inner Race Fault, and Outer Race Fault.

A balanced sample of 128 observations was analyzed, consisting of 32 observations per class. Equal class representation was retained to prevent overall accuracy from being dominated by one machine condition. Nine time-domain vibration features were used: maximum amplitude, minimum amplitude, mean, standard deviation, RMS, skewness, kurtosis, crest factor, and form factor. The use of these statistical indicators is consistent with feature-based rotating-machinery diagnosis, where vibration characteristics provide compact information on mechanical condition (Jorge et al., 2024).

Data Preprocessing

All nine input variables were inspected for missing and invalid values before model training. Numerical standardization was performed for algorithms sensitive to feature scale. The standardized value of observation i for feature j was calculated as:

$$z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \tag{1}$$

where x_{ij} represents the original value, μ_j is the mean of feature j calculated from the training data, and σ_j is its training-data standard deviation. Parameters estimated from the test observations were not used during preprocessing.

The data were initially divided using a stratified 75:25 training-test split with random seed 42. This produced 96 training observations and 32 held-out testing observations. The fixed split provided an initial diagnostic assessment, while repeated cross-validation was used as the primary performance evaluation because it provides a more stable estimate across multiple data partitions.

Classification Models

The principal deep learning model was a Deep Multilayer Perceptron with nine input neurons, followed by hidden layers containing 64, 32, and 16 neurons. ReLU activation was used in the hidden layers, while the output layer contained four neurons corresponding to the four bearing conditions. The architecture can be summarized as 9 → 64 → 32 → 16 → 4.

The output probabilities were generated using the Softmax function:

$$P(y = k \mid \mathbf{x}) = \frac{e^{a_k}}{\sum_{j=1}^{K} e^{a_j}} \tag{2}$$

where a_k is the output score for class k, and $K = 4$ is the number of machine-condition classes. Categorical cross-entropy was used as the classification loss:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} y_{ik} \ln(\hat{p}_{ik}) \tag{3}$$

where N is the number of observations, y_{ik} is the true class indicator, and $\hat{p}_{ik}$ is the predicted probability for observation i in class k.

Two conventional classifiers were included as benchmarks. SVM used the radial basis function kernel to model nonlinear class boundaries. Random Forest used an ensemble of 500 decision trees. The same feature variables and validation partitions were applied to all three models to ensure that model comparisons were not confounded by different data inputs.

Validation Procedure

Repeated Stratified 5-Fold Cross-Validation was used as the primary model-evaluation procedure. Five folds were repeated 10 times, generating 50 evaluation folds for each classifier. Stratification maintained the distribution of the four machine-condition classes within each fold. Preprocessing parameters were estimated from the training portion of each fold and applied to the corresponding validation portion to minimize information leakage.

Repeated cross-validation was selected because performance estimated from one train-test partition can be unstable in a relatively small dataset. The procedure provides a distribution of performance scores rather than a single accuracy value. Prawin (2025) emphasizes the importance of evaluating diagnostic stability rather than relying exclusively on overall accuracy, particularly when assessing fault-classification systems.

Performance Metrics

Model performance was evaluated using accuracy, macro precision, macro recall, macro F1-score, and confusion matrices. Multiclass accuracy was calculated as:

$$\text{Accuracy} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}\,(y_i = \hat{y}_i) \tag{4}$$

where N is the total number of observations and I(·) equals 1 when the predicted class is correct and 0 otherwise. For class k, precision was calculated as:

$$\text{Precision}_k = \frac{TP_k}{TP_k + FP_k} \tag{5}$$

Recall was calculated as:

$$\text{Recall}_k = \frac{TP_k}{TP_k + FN_k} \tag{6}$$

The class-specific F1-score was calculated as:

$$F1_k = 2\frac{\text{Precision}_k \times \text{Recall}_k}{\text{Precision}_k + \text{Recall}_k} \tag{7}$$

Macro F1-score assigned equal importance to every class:

$$\text{Macro-F1} = \frac{1}{K}\sum_{k=1}^{K} F\,1_k \tag{8}$$

where $K = 4$. Macro averaging was selected because it prevents strong performance in one class from dominating interpretation of the overall diagnostic result.

Statistical Analysis

Descriptive statistics were first used to examine differences among vibration characteristics across the four bearing conditions. Model performance was then summarized as the mean and standard deviation across the 50 repeated cross-validation folds. A 95% confidence interval was calculated for the mean performance of the Deep MLP.

H1 was evaluated using Deep MLP accuracy, macro precision, macro recall, macro F1-score, and cross-validation stability. H2 was evaluated by comparing paired macro F1-scores obtained from identical folds for Deep MLP, SVM-RBF, and Random Forest.

Because several performance scores approached the upper boundary of 1.00 and the paired differences did not satisfy a conventional normal-distribution assumption, the Wilcoxon signed-rank test was used for pairwise model comparison. Statistical significance was established at $\alpha = 0.05$, and a result was considered statistically significant when $p < 0.05$. All analyses were conducted using a consistent computational environment and fixed randomization settings to improve reproducibility.

Results and Discussion

Characteristics of the Vibration Data

The analysis used vibration-derived features from the Case Western Reserve University bearing dataset. Four operating conditions were evaluated: Normal, Ball Fault, Inner Race Fault, and Outer Race Fault. The fault classes represented a defect diameter of 0.007 inch at a motor load of 1 hp. The analytical dataset contained 128 balanced observations, with 32 observations assigned to each machine condition.

Nine time-domain vibration characteristics were analyzed, including maximum amplitude, minimum amplitude, mean, standard deviation, RMS, skewness, kurtosis, crest factor, and form factor. Descriptive analysis showed clear differences in vibration intensity among the four conditions. Outer Race Fault generated the largest RMS and standard deviation, whereas the normal condition produced the lowest vibration magnitude. Table 1 summarizes the principal descriptive statistics.

Table 1. Descriptive Statistics of the Main Vibration Features

Machine Condition	Mean SD	Mean RMS	Mean Kurtosis	Mean Crest Factor
Normal	0.0647	0.0656	-0.1257	2.9986
Ball Fault	0.1378	0.1391	0.0043	3.2577
Inner Race Fault	0.2800	0.2808	4.4253	5.1541
Outer Race Fault	1.0598	1.0597	3.8034	4.5408

Source: *Authors' analysis.*

The increase in RMS from 0.0656 under normal conditions to 1.0597 under Outer Race Fault indicates substantial differences in vibration energy. Inner Race Fault also produced higher kurtosis than the other classes, suggesting stronger impulsive behavior. These variations provide a physical basis for the subsequent classification analysis. Su et al. (2024a) similarly explain that changes in vibration amplitude and statistical structure are useful indicators for data-driven fault diagnosis.

Deep Learning Classification Performance

The Deep MLP was trained using standardized vibration features. The architecture consisted of three hidden layers with 64, 32, and 16 neurons, respectively. The data were initially divided using a stratified training-test procedure to preserve the proportion of each fault class.

On the fixed test partition, the model correctly classified all test observations. However, this result was not treated as the main indicator of model performance because a single partition can produce an optimistic estimate when classes are strongly separable. Therefore, repeated stratified cross-validation was used as the principal evaluation procedure.

The repeated 5-fold cross-validation with 10 repetitions produced a mean classification accuracy of 99.53% and a mean macro F1-score of 99.53%. The standard deviation of accuracy was 1.28 percentage points, indicating limited but observable variation among the evaluation folds. Table 2 reports the complete Deep MLP performance summary.

Table 2. Deep MLP Classification Performance

Evaluation Metric	Result
Mean Accuracy	99.53%
SD of Accuracy	1.28%
Mean Macro Precision	99.60%
Mean Macro Recall	99.52%
Mean Macro F1-score	99.53%

Evaluation Metric	Result
SD of Macro F1-score	1.30%
95% CI of Mean Accuracy	99.17%-99.90%

Source: *Authors' analysis based on repeated stratified 5-fold cross-validation.*

The model achieved perfect classification in 44 of the 50 evaluation folds. In six folds, accuracy decreased to approximately 96%. The presence of these errors is important because it shows that the classes were highly separable but not perfectly distinguishable under every resampling condition. The finding supports H1, which proposed that vibration information contains discriminative characteristics that allow machine conditions to be classified using deep learning. Nevertheless, the result should be interpreted within the characteristics of the dataset. Matania et al. (2024) caution that diagnostic accuracy obtained from controlled benchmark datasets may be substantially higher than performance obtained from naturally developing industrial faults.

Comparison with Conventional Machine-Learning Models

The Deep MLP was compared with Support Vector Machine using a radial basis function kernel and Random Forest. The same data partitions were used for all models to provide a consistent evaluation. Table 3 shows that Deep MLP performed extremely well but did not exceed the conventional classifiers.

Table 3. Comparison of Classification Models

Model	Mean Accuracy (%)	Mean Macro F1-score (%)
Deep MLP	99.53	99.53
SVM-RBF	100.00	100.00
Random Forest	100.00	100.00

Source: *Authors' analysis based on repeated stratified 5-fold cross-validation.*

The difference between Deep MLP and the two benchmark models was approximately 0.47 percentage points in macro F1-score. A Wilcoxon signed-rank comparison produced $p = 0.0256$ for Deep MLP versus SVM-RBF and the same probability value for Deep MLP versus Random Forest. Table 4 summarizes the paired statistical comparison.

Table 4. Statistical Comparison of Model Performance

Comparison	Difference in Macro F1	p-value	Decision
Deep MLP vs. SVM-RBF	-0.47 percentage points	0.0256	Significant
Deep MLP vs. Random Forest	-0.47 percentage points	0.0256	Significant

Source: *Authors' analysis.*

Based on these results, H2 was not supported. Deep learning did not significantly outperform the conventional machine-learning models. Instead, SVM-RBF and Random Forest produced slightly stronger performance. This finding provides a useful practical interpretation. Deep learning does not automatically outperform conventional algorithms when the input already consists of highly discriminative engineered features. Nguyen et al. (2025) also showed that the relative performance of machine-learning and deep-learning approaches depends strongly on the characteristics of the vibration representation.

The performance difference was nevertheless small. The Deep MLP macro F1-score of 99.53% was only 0.47 percentage points lower than the best classifiers. Therefore, the result does not indicate poor deep-learning performance. Rather, it indicates that additional model complexity provided no measurable advantage for the selected feature-based dataset.

Classification Stability

The repeated cross-validation results provide a more realistic assessment than the single test split. Although the Deep MLP obtained perfect classification on the initial held-out observations, its performance varied slightly across repeated partitions. Table 5 summarizes this stability across all 50 evaluation folds.

Table 5. Stability of Deep MLP across 50 Evaluation Folds

Performance Category	Number of Folds	Percentage of Folds
100% accuracy	44	88.0%
Approximately 96% accuracy	6	12.0%
Below 95% accuracy	0	0.0%
Total	50	100.0%

Source: *Authors' analysis.*

The model achieved perfect classification in 88% of evaluation folds, while small classification errors appeared in 12% of folds. No fold produced accuracy below 95%. This pattern indicates high diagnostic stability while avoiding the assumption that the model will always classify every observation correctly. The small variation across folds may result from borderline observations whose statistical vibration characteristics overlap between fault conditions. More detailed frequency or time-frequency information may improve separation for these difficult observations.

Interpretation for Predictive Maintenance

The results indicate that vibration-derived statistical characteristics provide sufficient information for reliable diagnosis under the tested condition. The mean repeated-validation accuracy of 99.53% suggests that the Deep MLP can identify bearing conditions consistently when operating conditions are similar to those represented during training.

However, predictive maintenance requires more than high classification accuracy. Industrial machinery operates under variations in load, rotational speed, temperature, structural vibration, and sensor noise. A model that performs strongly under one controlled operating condition may experience performance degradation when the data distribution changes. Liu et al. (2024) demonstrated that domain differences can substantially influence rotating-machinery fault diagnosis, which explains the increasing use of domain-adaptation methods.

The current finding should therefore be interpreted as evidence of fault discriminability under a controlled operating condition, rather than proof of universal industrial robustness. This distinction makes the result more relevant to practical predictive maintenance because it establishes what the model can reliably achieve while acknowledging what has not yet been tested.

Hypothesis Evaluation

The final hypothesis evaluation is presented in Table 6. The results clearly distinguish the ability of vibration features to support accurate diagnosis from the separate question of whether Deep MLP is superior to conventional classifiers.

Table 6. Summary of Hypothesis Testing

Hypothesis	Main Evidence	Decision
H1: Vibration features enable effective classification of normal and faulty bearing conditions	Accuracy = 99.53%; Macro F1 = 99.53%; 95% CI = 99.17%-99.90%	Supported
H2: Deep MLP outperforms SVM-RBF and Random Forest	Deep MLP = 99.53%; SVM-RBF = 100%; RF = 100%; p = 0.0256	Not Supported

Source: *Authors' analysis.*

The rejection of H2 is an important finding rather than a weakness of the study. It demonstrates that model complexity should be matched to the structure of the input data. When fault-sensitive statistical features already provide strong class separation, conventional machine-learning algorithms may achieve comparable or slightly superior performance at lower computational cost.

Discussion

The results demonstrate that vibration-based fault diagnosis can achieve highly reliable classification when fault-sensitive information is clearly represented in the input features. The Deep MLP achieved a mean macro F1-score of 99.53%, confirming that nonlinear neural learning can effectively separate Normal, Ball Fault, Inner Race Fault, and Outer Race Fault conditions. The repeated cross-validation procedure provides stronger evidence than a single train-test split because model performance was evaluated across 50 different partitions.

At the same time, the findings challenge the assumption that deep learning must always outperform conventional machine learning. Both SVM-RBF and Random Forest reached perfect repeated-validation performance, whereas Deep MLP exhibited a small number of classification errors. This finding can be explained by the nature of the input. Statistical features such as RMS and standard deviation already provided substantial separation among the fault conditions. Consequently, sophisticated hierarchical feature learning offered limited additional benefit.

The descriptive findings reinforce this interpretation. Mean RMS increased from 0.0656 for Normal operation to 0.1391 for Ball Fault, 0.2808 for Inner Race Fault, and 1.0597 for Outer Race Fault. Such pronounced differences allow relatively simple nonlinear decision boundaries to distinguish the classes. Deep learning becomes more valuable when the input consists of raw high-dimensional signals for which discriminative representations must be learned automatically.

The findings have practical implications for predictive maintenance. An industrial diagnostic system should use the simplest model capable of achieving the required sensitivity and reliability. A deep model may be appropriate when raw signals, multiple sensors, or complex operating environments are involved. Conversely, Random Forest or SVM can remain attractive when reliable statistical features have already been extracted because they require fewer computational resources and are easier to deploy.

Several limitations should be recognized. The current analysis uses a controlled CWRU-derived dataset and one operating load. The faults were experimentally introduced and therefore may be easier to distinguish than naturally developing industrial degradation. In addition, the analysis used time-domain features rather than the complete raw vibration waveform. The results should consequently not be generalized directly to all industrial rotating machinery. Future experiments should extend the analysis across multiple load and speed conditions, incorporate realistic noise, and compare raw time-domain signals with STFT or wavelet representations. Cross-load testing is especially important because it prevents models from relying on operating-condition-specific patterns.

Conclusion

This study evaluated vibration-based bearing fault diagnosis using nine time-domain statistical features and compared the performance of Deep MLP, SVM-RBF, and Random Forest under identical validation procedures. The Deep MLP achieved a mean accuracy and macro F1-score of 99.53% across repeated stratified 5-fold cross-validation, confirming that the selected vibration features provide strong discriminative information for distinguishing Normal, Ball Fault, Inner Race Fault, and Outer Race Fault conditions. Accordingly, H1 was supported. However, SVM-RBF and Random Forest both achieved 100.00% mean accuracy and macro F1-score, while the comparison with Deep MLP produced $p = 0.0256$. Therefore, H2 was not supported. These findings indicate that deep learning does not automatically outperform conventional machine-learning methods when the input already contains highly discriminative engineered vibration features.

The results have practical implications for predictive maintenance because model selection should consider diagnostic reliability, computational complexity, and the characteristics of available sensor data rather than model sophistication alone. SVM-RBF and Random Forest provide efficient alternatives for feature-based diagnosis, whereas deep learning may offer greater value when models must learn representations directly from raw, high-dimensional, or multi-sensor vibration signals. The study is limited to a controlled CWRU-derived dataset, one operating load, and experimentally introduced bearing faults. Future research should therefore evaluate cross-load and cross-speed generalization, realistic noise conditions, naturally developing industrial faults, raw-signal and time-frequency representations, multi-sensor fusion, and remaining useful life estimation to strengthen the transition from fault classification toward comprehensive predictive-maintenance systems.

Acknowledgements

The authors gratefully acknowledge the Case Western Reserve University Bearing Data Center for providing the publicly accessible bearing vibration data underlying this study. The authors also appreciate the contributions of researchers whose previous studies provided methodological and theoretical foundations for vibration-based fault diagnosis and predictive maintenance.

Author Contributions

Conceptualization: Aditya Bagus Santoso; Rina Oktavia Sari
Data curation: Rina Oktavia Sari; Rafi Alfarizi Hidayat
Formal analysis: Aditya Bagus Santoso; Rafi Alfarizi Hidayat
Investigation: Aditya Bagus Santoso; Rina Oktavia Sari; Rafi Alfarizi Hidayat
Methodology: Aditya Bagus Santoso; Rina Oktavia Sari
Project administration: Aditya Bagus Santoso
Supervision: Maya Fitri Handayani
Validation: Rafi Alfarizi Hidayat; Maya Fitri Handayani
Visualization: Rafi Alfarizi Hidayat
Writing - original draft: Aditya Bagus Santoso; Rina Oktavia Sari
Writing - review and editing: Aditya Bagus Santoso; Rina Oktavia Sari; Rafi Alfarizi Hidayat; Maya Fitri Handayani

Funding Statement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of Interest

The authors declare that they have no conflict of interest related to this manuscript.

Statement on the Use of Artificial Intelligence

The authors used ChatGPT by OpenAI and Claude AI by Anthropic to support language refinement, academic editing, structural organization, and improvement of manuscript clarity. These tools were not listed as authors and did not replace the authors' responsibility for the research design, data interpretation, accuracy, originality, citation verification, confidentiality, research ethics, or final scientific content. All AI-assisted output was reviewed and verified by the authors, who take full responsibility for the final manuscript.

References

Ahsan, M., Hassan, M. W., Rodríguez, J., & Abdelrahem, M. (2025). Enhanced fault diagnosis in rotating machinery using a hybrid CWT-LeNet-5-LSTM model: Performance across various load conditions. IEEE Access, 13, 1026–1045. https://doi.org/10.1109/ACCESS.2024.3522948

Ali, M. I., Lai, N. S., & Abdulla, R. (2024). Predictive maintenance of rotational machinery using deep learning. International Journal of Electrical and Computer Engineering, 14(1), 1112–1121. https://doi.org/10.11591/ijece.v14i1.pp1112-1121

Almutairi, K. M., Sinha, J. K., & Wen, H. (2024). Fault detection of rotating machines using poly-coherent composite spectrum of measured vibration responses with machine learning. Machines, 12(8), 573. https://doi.org/10.3390/machines12080573

Anshori, M. I., & Dhini, A. (2025). Fault detection of rotating machinery in the petrochemical industry using a deep learning based approach: TabNet-WGAN. Eastern-European Journal of Enterprise Technologies, 3(1[135]), 90–99. https://doi.org/10.15587/1729-4061.2025.332597

Atmaji, F. T. D., Jamasri, Yuniarto, H. A., & Miasa, I. M. (2025). Experimental investigation of shaft misalignment effects on bearing reliability through vibration signal analysis using machine learning and deep learning. Results in Engineering, 27, 106754. https://doi.org/10.1016/j.rineng.2025.106754

Bagri, I., Tahiry, K., Hraiba, A., Touil, A., & Mousrij, A. (2024). Vibration signal analysis for intelligent rotating machinery diagnosis and prognosis: A comprehensive systematic literature review. Vibration, 7(4), 1013–1062. https://doi.org/10.3390/vibration7040054

Das, O., Das, D. B., & Birant, D. (2023). Machine learning for fault analysis in rotating machinery: A comprehensive review. Heliyon, 9(6), e17584. https://doi.org/10.1016/j.heliyon.2023.e17584

Dong, Z., Zhao, D., & Cui, L. (2024). An intelligent bearing fault diagnosis framework: One-dimensional improved self-attention-enhanced CNN and empirical wavelet transform. Nonlinear Dynamics, 112, 6439–6459. https://doi.org/10.1007/s11071-024-09389-y

Gao, W., Fu, D., Yu, J., & Akoudad, Y. (2025). HDNAT: A Transformer-based fault diagnosis model for rotating machinery across varying damage degrees. IEEE Transactions on Instrumentation and Measurement, 74, 1–16. https://doi.org/10.1109/TIM.2025.3623791

Hu, J. (2025). Intelligent fault diagnosis method for rolling bearings based on SSA-CNN-Transformer. Highlights in Science, Engineering and Technology, 143, 107–118. https://doi.org/10.54097/qfj5xs82

Jiang, W., Zhou, R., & Jia, H. (2024). Research on mechanical equipment fault diagnosis and prediction technology based on vibration signal analysis. In 2024 International Conference on Power, Electrical Engineering, Electronics and Control (PEEEC) (pp. 494–498). IEEE. https://doi.org/10.1109/PEEEC63877.2024.00096

Jiang, Z., Liu, D., & Cui, L. (2024). Deep adaptively dynamic edge graph convolution network with attention weight and high-dimension affinity feature graph for rotating machinery fault diagnosis. Measurement Science and Technology. https://doi.org/10.1088/1361-6501/ad9e1d

Jorge, A. R. F., Gouveia, E. B., da Cunha, M. J. R., Cavallini, A. A., Jr., & Freitas, L. C. G. (2024). Rotodynamics multi-fault diagnosis through time domain parameter analysis with MLP: A comprehensive study. In 2024 International Workshop on Artificial Intelligence and Machine Learning for Energy Transformation (AIE). IEEE. https://doi.org/10.1109/AIE61866.2024.10561397

Jweri, A.-R. K., Ogaili, A. A. F., Amin, S. A., Khalaf, M. I., Al-Haddad, L. A., Jaber, A. A., & Karkhi, M. I. (2025). Enhancing predictive maintenance in energy systems using a hybrid Kolmogorov-Arnold Network with Short-Time Fourier Transform framework for rotating machinery. ASEAN Journal of Science and Engineering, 5(2), 465–494. https://doi.org/10.17509/ajse.v5i2.89023

Kang, J., Luo, Y., Wang, P., Wei, Y., & Zhou, Y. (2023). Fault diagnosis of rotating machinery under complex conditions based on multi-scale convolutional neural networks. Journal of Physics: Conference Series, 2658(1), 012038. https://doi.org/10.1088/1742-6596/2658/1/012038

Kibrete, F., Woldemichael, D. E., & Gebremedhen, H. S. (2025). Fault diagnosis of rotating machines based on combination of one-dimensional convolutional neural network and long short-term memory in variable working conditions. Journal of Engineering, 2025, 1670810. https://doi.org/10.1155/je/1670810

Li, J., & Bai, M. (2024). Review of research on signal decomposition and fault diagnosis of rolling bearing based on vibration signal. Measurement Science and Technology. https://doi.org/10.1088/1361-6501/ad4eff

Li, X., Tang, J., Zhang, J., Pang, P., & Hou, Y. (2025). An improved CS-Transformer for fault diagnosis of rotating machinery bearings under strong noise conditions. Latin American Journal of Solids and Structures, 22(9), e8697. https://doi.org/10.1590/1679-7825/e8697

Liu, D., Cui, L., & Cheng, W. (2024). Interpretable domain adaptation transformer: A transfer learning method for fault diagnosis of rotating machinery. Structural Health Monitoring. https://doi.org/10.1177/14759217241249656

Ma, J., Huang, J., Liu, S., Luo, J., & Jing, L. (2024). A self-attention Legendre graph convolution network for rotating machinery fault diagnosis. Sensors, 24(17), 5475. https://doi.org/10.3390/s24175475

Matania, O., Dattner, I., Bortman, J., Kenett, R. S., & Parmet, Y. (2024). A systematic literature review of deep learning for vibration-based fault diagnosis of critical rotating machinery: Limitations and challenges. Journal of Sound and Vibration, 590, 118562. https://doi.org/10.1016/j.jsv.2024.118562

Mayo, S. A., Rehman, S., & Cai, Z. (2024). High-accuracy gearbox fault detection using deep learning on vibrational data. Journal of Physics: Conference Series, 2853(1), 012066. https://doi.org/10.1088/1742-6596/2853/1/012066

Mrabti, A., Younes, R., Ouelaa, N., Kebabsa, T., & Ouelaa, Z. (2025). Robust fault detection and severity classification in rotating machinery using VMD-LSTM for limited data scenarios. Advances in Mechanical Engineering, 17. https://doi.org/10.1177/16878132251342909

Nguyen, T. H. T., Pham, N. V., & Hoang, Q. V. (2025). Bearing fault diagnosis by machine learning and deep learning-based models: A comparative study applying for HUST bearing dataset. Journal of Research on Army Science and Technology, 103, 31–39. https://doi.org/10.54939/1859-1043.j.mst.103.2025.31-39

Prawin, J. (2025). Deep learning neural networks with input processing for vibration-based bearing fault diagnosis under imbalanced data conditions. Structural Health Monitoring, 24(2), 883–908. https://doi.org/10.1177/14759217241246508

Rezazadeh, N., De Oliveira, M., Lamanna, G., Perfetto, D., & De Luca, A. (2025). WaveCORAL-DCCA: A scalable solution for rotor fault diagnosis across operational variabilities. Electronics, 14(15), 3146. https://doi.org/10.3390/electronics14153146

Siavash-Abkenari, N., Rahmani-Sane, G., Torkaman, H., & Alipoor, G. (2024). Exploring a cutting-edge framework for bearing fault detection: A synergistic approach integrating statistical analysis and deep learning methods. IEEE Journal of Emerging and Selected Topics in Industrial Electronics. https://doi.org/10.1109/JESTIE.2024.3373313

Song, B., Liu, Y., Fang, J., Liu, W., Zhong, M., & Liu, X. (2024). An optimized CNN-BiLSTM network for bearing fault diagnosis under multiple working conditions with limited training samples. Neurocomputing, 574, 127284. https://doi.org/10.1016/j.neucom.2024.127284

Su, H., Xiang, L., & Hu, A. (2024). Application of deep learning to fault diagnosis of rotating machineries. Measurement Science and Technology, 35(4), 042003. https://doi.org/10.1088/1361-6501/ad1e20

Su, N., Chen, Y., Liu, Y., Zhang, Q., Zhou, L., He, Y., & Chang, X. (2024). Fault diagnosis of rotating machinery under variable operating conditions based on multi-feature and transfer learning. In 2024 6th International Conference on Internet of Things, Automation and Artificial Intelligence (IoTAAI) (pp. 190–194). IEEE. https://doi.org/10.1109/IOTAAI62601.2024.10692373

Tran, K., Vu, H., Pham, L., Boudaoud, N., & Nguyen, H.-S.-H. (2024). Robust-MBDL: A robust multi-branch deep-learning-based model for remaining useful life prediction of rotating machines. Mathematics, 12(10), 1569. https://doi.org/10.3390/math12101569

Wang, S., Su, X. L., Li, J., Li, F., Li, M., Ren, Y., Wang, G., Shi, N., & Huang, Q. (2025). A synergistic fault diagnosis method for rolling bearings: Variational mode decomposition coupled with deep learning. Electronics, 14(18), 3714. https://doi.org/10.3390/electronics14183714

Wang, X., Zhou, W., & Li, X. (2023). Bearing fault diagnosis based on graph formulation and graph convolutional network. Journal of Dynamics, Monitoring and Diagnostics, 2(4), 252–261. https://doi.org/10.37965/jdmd.2023.468

Wei, L., Peng, X., & Cao, Y. (2024). Multi-scale graph Transformer for rolling bearing fault diagnosis. Eksploatacja i Niezawodność – Maintenance and Reliability. https://doi.org/10.17531/ein/194779

Xie, F., Wang, G., Shang, J., Sun, E., & Xie, S. (2023). Gearbox fault diagnosis based on multi-sensor deep spatiotemporal feature representation. Mathematics, 11(12), 2679. https://doi.org/10.3390/math11122679

Yang, Z.-Z., Li, W., Yuan, F., Zhi, H., Guo, M., Xin, B., & Gao, Z. (2025). Hybrid CNN-BiLSTM-MHSA model for accurate fault diagnosis of rotor motor bearings. Mathematics, 13(3), 334. https://doi.org/10.3390/math13030334

Zhang, J., Zhang, Q., Feng, W., Qin, X., & Sun, Y. (2023). Gearbox fault diagnosis based on frequency-domain Gramian angular difference field and deep convolutional neural network. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 237(21), 5187–5202. https://doi.org/10.1177/09544062231157189

Zhang, Q., & Deng, L. (2023). An intelligent fault diagnosis method of rolling bearings based on short-time Fourier transform and convolutional neural network. Journal of Failure Analysis and Prevention, 23, 795–811. https://doi.org/10.1007/s11668-023-01616-9

Zhang, Q., Wei, X., Wang, Y., & Hou, C. (2024). Convolutional neural network with attention mechanism and visual vibration signal analysis for bearing fault diagnosis. Sensors, 24(6), 1831. https://doi.org/10.3390/s24061831

Zhang, S., Zhou, J., Ma, X., Pirttikangas, S., & Yang, C. (2024). TSViT: A time series vision transformer for fault diagnosis of rotating machinery. Applied Sciences, 14(23), 10781. https://doi.org/10.3390/app142310781