DOI 10.52209/2706-977X_2026_2_46 **IRSTI 55.01.37** **UDC 004.50**

Fault Diagnosis Using Limited Vibration Dataset with Hyperparameter Optimization

Zarog M.*, Al-Adawi S., Al-Nabhani M., Al-Aufi I.
Department of Mechanical and Industrial Engineering, College of Engineering, Sultan Qaboos University
*corresponding author

Abstract. Fault diagnosis in mechanical systems plays a critical role in predictive maintenance and operational safety. This study presents a machine learning-based approach for identifying common mechanical faults using a limited vibration dataset. An experimental setup was developed to simulate and record vibrations from various conditions, including healthy operation, bearing faults, gear defects, and screw looseness. Vibration data were collected in three dimensions and transformed into the frequency domain. Several classification algorithms, including Naïve Bayes, Logistic Regression, Decision Table, Random Tree, and Random Forest, were trained to detect these faults. Additionally, Auto-WEKA was employed to optimize both algorithm selection and hyperparameters. Among all tested methods, the Random Forest classifier with hyperparameter optimization achieved superior performance with a classification accuracy of 97.1% and an estimated error rate of 0.85%. The study highlights that even with a relatively small dataset, high diagnostic performance is achievable through proper algorithm selection and hyperparameter tuning. The findings are particularly relevant for applications in industrial environments where collecting large labelled datasets is costly or impractical.

Keywords: condition monitoring, machine learning, fault diagnosis, vibration monitoring, preventive maintenance, and fault detection

Introduction

To keep rotating machinery, such as pumps and motors, reliable and functionally efficient, fault detection and classification are crucial. Vibration analysis has long been used as an important method for monitoring the status of such equipment. Machine learning techniques can be utilized to analyse measured vibration data. However, these approaches highly depend on past historical data for training, which might not be available, or costly. Traditional machine learning methods, though effective in many cases, often require large amounts of labeled data to achieve high accuracy in fault classification. This limitation has driven researchers to explore advanced machine learning techniques, such as Hyperparameter Optimization This adaptability is critical in industrial settings where the amount of data available for model training may be limited. Several studies have shown that ANNs can outperform traditional models in terms of classification accuracy when dealing with small datasets. For instance, Kankar et al. [1] demonstrated the effectiveness of ANNs in classifying bearing faults using a relatively small vibration dataset, highlighting the model's ability to generalize from limited information. Similarly, Khan and Yairi [2] investigated fault diagnosis using deep learning approaches and found that neural networks could achieve superior results with constrained data availability. Research by Jia et al. [3] demonstrated that ANN-based models could automatically identify relevant features from raw vibration signals, and had shown promise in limited data scenarios, their success largely depends on network architecture, training techniques, and data preprocessing. Also, Lei et al. [4] proposed an improved ANN framework for fault classification in mechanical systems, emphasizing the importance of these techniques in optimizing model performance on small datasets.

The use of convolutional neural networks (CNNs) has also been explored for fault classification. Bouvrie [5] and Chen et al. [6] showed that CNNs are capable of learning spatial hierarchies from vibration signals, enabling accurate classification of gearbox and motor faults. Gude et al. [7] provided evidence that time-frequency analysis combined with machine learning can detect engine-related anomalies effectively using vibration data. Many more attempts were carried out for fault discriminative features using unlabelled and a limited number of labelled samples for classification [8-12]. Deep learning was not pursued because the available dataset was relatively small, increasing the risk of severe overfitting and poor generalization. Traditional machine learning methods were considered more appropriate because they typically perform better than deep learning on limited datasets with structured features. Considering the risk of overfitting and the absence of large-scale training data, traditional machine learning models optimized through Auto-WEKA offered a more reliable and practical solution.

Despite these advances, many models require large datasets and computational resources. In this context, the current study emphasizes the practicality of using limited datasets and traditional classifiers, enhanced through Auto-WEKA optimization, to achieve comparable performance in fault diagnosis tasks. Given the increasing use of fault classification in industrial applications, it is essential to assess their efficiency, particularly when dealing with limited vibration datasets. This paper aims to explore the performance of hyperparameter optimization in classifying faults from vibration data, focusing on how the training methods can be optimized to ensure reliable fault detection with minimal data.

1. Experimental set up and data collection

A mechanical system, as shown in Figure.1, was designed to simulate different faults in the system components (e.g. bearings, gears and screws looseness). The electric motor provides the system with kinetic energy, through which the system is operated. The motor was installed in the base using screw and nuts. the power produced by the motor is transmitted through gear box, that alter the speed and torque, along the pulleys and belts to shaft, through bearing. The system was mounted on base frame via screws, as shown in Figure.1. Vibration was measured at different faulty scenarios as well as during normal condition (vibration when no fault is there). Measurements were taken for the vibration amplitude and frequency in the x, y, and z directions. Measurement was repeated 10 times for each condition (healthy condition, faulty bearings, faulty gears, and screws looseness). Figure.2 shows faults in gears, bearing, and looseness of the screw. Figure.3, 4, 5, and 6 shows vibration measurement in frequency domain for healthy and faulty components. The experimental readings from the sensor were taken over wide range of frequencies (0 to 512 Hz, with interval of 0.5 Hz). The readings were taken for the normal condition and then for different faults of each component.

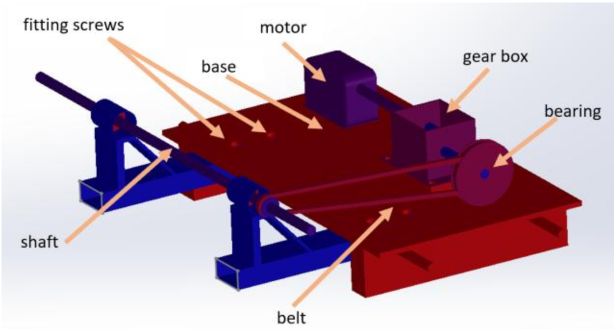

Fig.1. - 3D view of the system components.

(a) (b) (c)

Fig.2. - Faults in gears (a), bearing (b), and looseness (c)

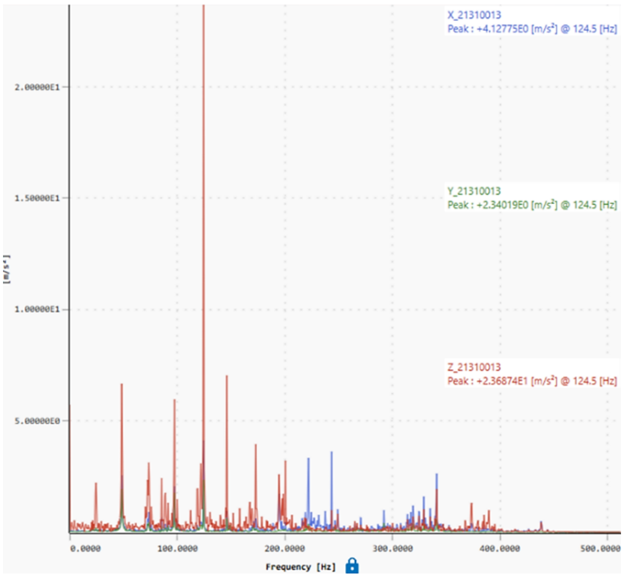

Fig.3. - Measured data of healthy system in frequency domain

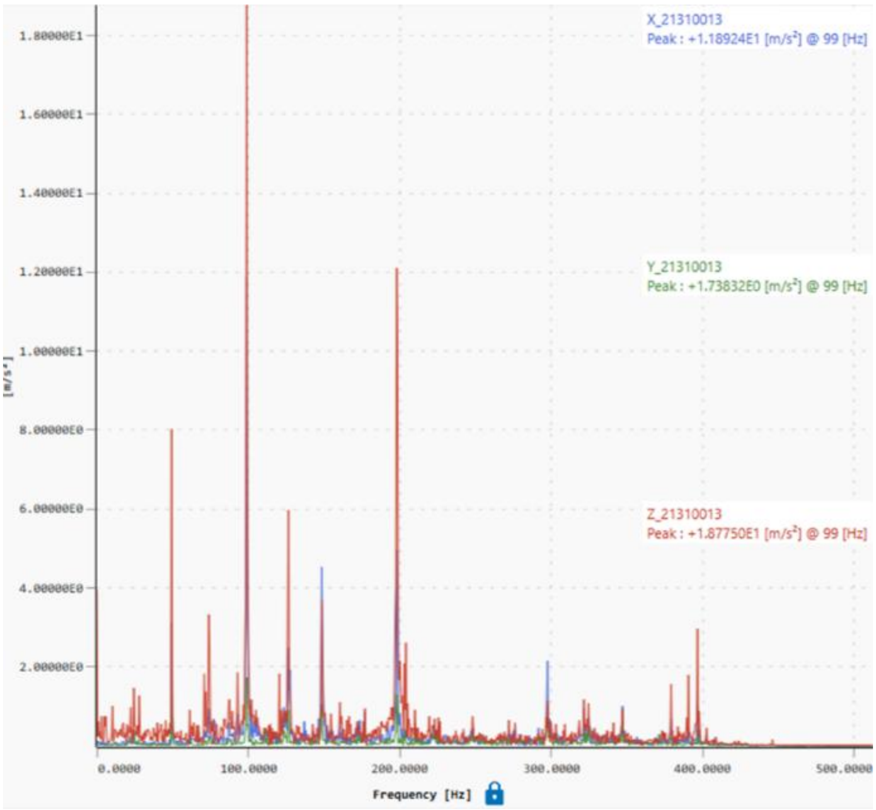

Fig.4. - Measured data with faulty system in frequency domain (Screw looseness fault)

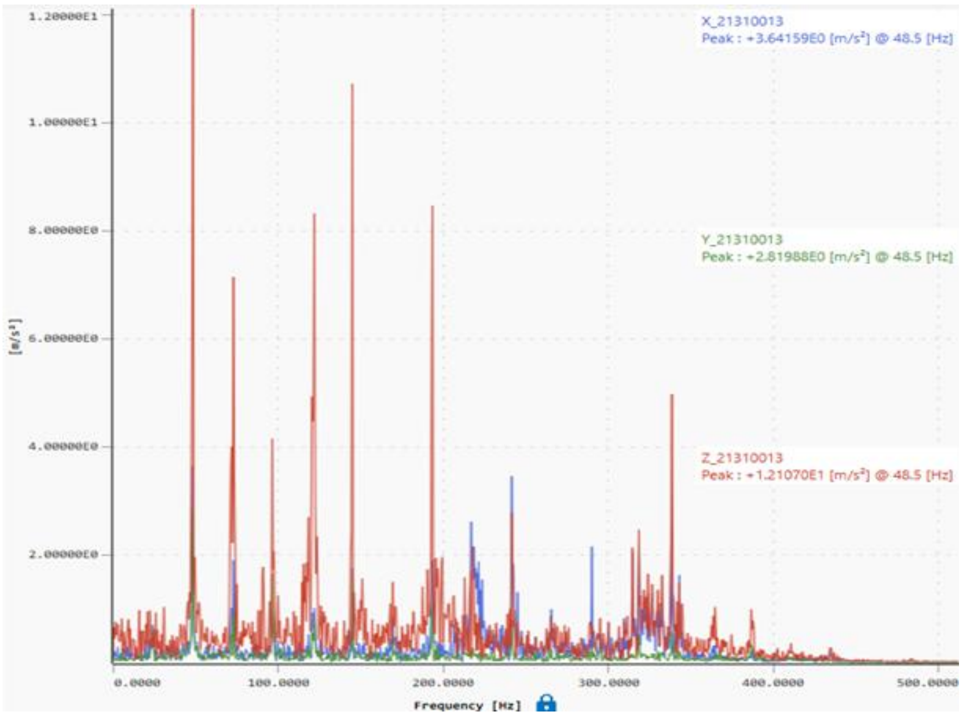

Fig.5. - Measured data with faulty system in frequency domain (Bearing fault)

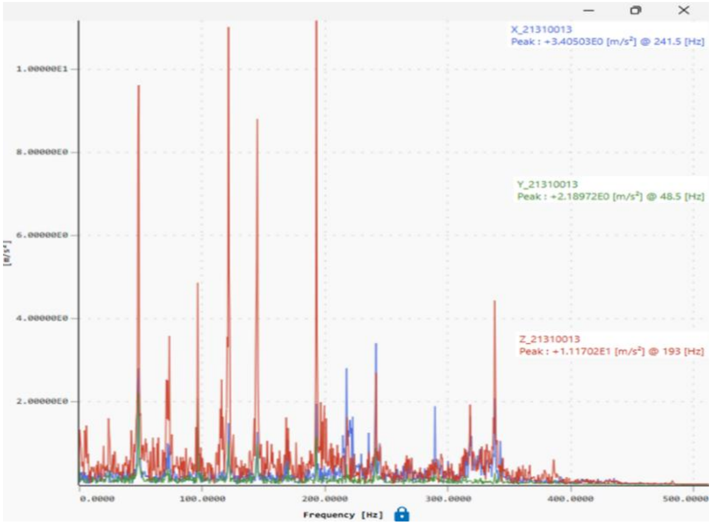

Fig.6. - Measured data with faulty system in frequency domain (Gear fault)

2. Results

The obtained dataset from the captured vibration signals were utilized for training using ML algorithms. The evaluation was conducted using five widely used classifiers - Naïve Bayes, Logistic Regression, Decision Table, Random Tree, and Random Forest - under two classification scenarios: multiclass classification involving four distinct fault conditions as shown in Table.1, and binary classification comparing normal operation against each individual fault type as shown in Table.2. The vibration signal for each of the four fault conditions was measured over the frequency range of 0–512 Hz with a frequency resolution of 0.5 Hz, resulting in 1024 readings for each fault condition. Since four fault conditions were considered, the total number of instances in the dataset was 4096.

Table 1. Classification results for all four conditions

Algorithm/Measures (using 10 folds)	Naïve Bayes	Logistic	Decision Table	Random Tree	Radom Forest (4 conditions)
Correctly Classified Instances	1150	1481	1701	1857	2163
Incorrectly Classified Instances	2950	2619	2399	2243	1937
correctly classified instances	28%	36.0%	42.0%	45.0%	53.0%
Mean absolute error	0.37	0.36	0.336	0.27	0.28
Root mean squared error	0.45	0.43	0.4101	0.52	0.38
Relative absolute error	98.0 %	97.0 %	89.6 %	72.9 %	74.6 %
Root relative squared error	104.0 %	98.4 %	94.7 %	120.8 %	88.6 %
Total Number of Instances	4096	4096	4096	4096	4096

Table 2. Classification results for normal conditions Vs other conditions

Class	Normal Vs looseness fault	Normal Vs bearing fault	Normal Vs Gear fault
Naïve Bayes			
Correctly Classified Instances	1068	1078	1178
Incorrectly Classified Instances	982	972	872
correctly classified instances %	52 %	52.6 %	57.5 %
Mean absolute error	0.49	0.48	0.44
Root mean squared error	0.54	0.55	0.51
Relative absolute error	97.1 %	96.1 %	88.9 %
Root relative squared error	108.0%	111.0 %	102.4 %
Total Number of Instances	2048	2048	2048

Continuation of Table 2

Logistic			
Correctly Classified Instances	1318	1224	1424
Incorrectly Classified Instances	732	826	626
correctly classified instances %	64.3 %	59.7 %	69.4 %
Mean absolute error	0.45	0.48	0.41
Root mean squared error	0.47	0.49	0.45
Relative absolute error	90.3 %	96.6 %	82.9 %
Root relative squared error	94.9 %	98.3 %	90.4 %
Total Number of Instances	2048	2048	2048
Decision Table			
Correctly Classified Instances	1365	1304	1492
Incorrectly Classified Instances	685	746	558
correctly classified instances %	66.6%	63.6 %	72.8 %
Mean absolute error	0.43	0.43	0.37
Root mean squared error	0.45	0.47	0.43
Relative absolute error	86.0 %	86.3 %	74.4 %
Root relative squared error	93.0 %	93.1 %	86.9 %
Total Number of Instances	2048	2048	2048
Random Tree			
Correctly Classified Instances	1478	1333	1552
Incorrectly Classified Instances	572	717	498
correctly classified instances %	72.1%	65.0 %	75.7 %
Mean absolute error	0.28	0.35	0.24
Root mean squared error	0.52	0.59	0.49
Relative absolute error	55.8 %	70.0 %	48.6 %
Root relative squared error	105.6%	118.3 %	98.6 %
Total Number of Instances	2048	2048	2048
Radom Forest algorithm			
Correctly Classified Instances	1596	1452	1674
Incorrectly Classified Instances	454	598	376
correctly classified instances %	77.9 %	70.8 %	81.7 %
Mean absolute error	0.30	0.35	0.26
Root mean squared error	0.39	0.44	0.37
Relative absolute error	60.3 %	70.9 %	51.2 %
Root relative squared error	78.2 %	87.1 %	73.4 %
Total Number of Instances	2048	2048	2048

3. Hyperparameter Optimization

Given a set of learning algorithms $\mathcal{A}$ and a limited amount of training data $D = \{(X_1, Y_1), \ldots\ldots, (X_n, Y_n)\}$, the goal of model selection is to determine the algorithm $A^* \in \mathcal{A}$ with optimal generalization performance. Generalization performance is estimated by splitting D into disjoint training and validation sets $D_{train}^{(i)}$ and $D_{valid}^{(i)}$, learning functions f_i by applying A^* to $D_{train}^{(i)}$, and evaluating the predictive performance of these functions on $D_{valid}^{(i)}$. This allows for the model selection problem to be written as [13]:

$$A^* \in argmine_{A^* \in \mathcal{A}} \; \frac{1}{k} \sum_{i=1}^{k} \mathcal{L}(A, D_{train}^{(i)}, D_{valid}^{(i)}).$$

where $\mathcal{L}(A, D_{train}^{(i)}, D_{valid}^{(i)})$ is the loss (here: misclassification rate) achieved by A When trained on $D_{train}^{(i)}$ and evaluated on $D_{valid}^{(i)}$.

We use k-fold cross-validation [18], which splits the training data into k equal-sized partitions $D_{valid}^{(1)}, \ldots,$ $D_{valid}^{(K)}$, and sets $D_{train}^{(i)} = D/D_{valid}^{(i)}$ for $i = 1, \ldots, K$.

More formally, following [15], we say that a hyperparameter λ_i is conditional on another hyperparameter λ_j, if λ_i is only active if the hyperparameter λ_j takes values from a given set $V_i(j) \subsetneq \Lambda_j$; In this case, we call λ_j a parent of λ_i. Conditional hyperparameters can in turn be parents of other conditional hyperparameters, giving rise to a tree-structured space [3] or, in some cases, a directed acyclic graph (DAG) [15]. Given such a structured space Λ, the (hierarchical) hyperparameter optimization problem can be written as [13]:

$$\lambda^* \in argmine_{\lambda \in \mathcal{A}} \ \frac{1}{k}\sum_{i=1}^{k} \mathcal{L}\left(A_\lambda, D_{train}^{(i)}, D_{valid}^{(i)}\right).$$

Given a set of algorithms $\mathcal{A} = \{ A^{(1)}, \dots, A^{(k)} \}$ with associated hyperparameter spaces $\Lambda^{(1)}, \dots, \Lambda^{(k)}$, we define the combined algorithm selection and hyperparameter optimization problem (CASH) as computing [13]:

$$A^*{}_{\lambda^*} \in argmine_{A^{(j)} \in \mathcal{A},\ \lambda \in \Lambda^{(j)}} \ \frac{1}{k}\sum\nolimits_{i=1}^{k} \mathcal{L}\left(A_\lambda{}^{(j)}, D_{train}^{(i)}, D_{valid}^{(i)}\right). \qquad (1)$$

Auto-WEKA can automatically and simultaneously choose a learning algorithm and set its hyperparameters to optimize empirical performance [13]. Auto-WEKA tested various algorithms (e.g., SVMs, decision trees, etc.) and found that the Random Forest classifier produced the lowest estimated error rate on the dataset. Auto-WEKA optimized for the lowest error rate, which is defined as the percentage of incorrect predictions, which was found to be 0.0085. This means only about 0.85% of the test instances are expected to be misclassified using this configuration.The forest is composed of 130 decision trees. More trees can lead to better performance (to a point), but with increased computation. Maximum depth of trees: Limits the depth of each individual tree to prevent overfitting. A value of 15 allows moderately deep trees that capture complex patterns without going too deep. No feature selection (attribute search) was applied like BestFirst or GreedyStepwise.

Table 3. Classification results for all four conditions using hyperparameter optimization

Correctly Classified Instances	3982
Incorrectly Classified Instances	118
correctly classified instances %	97.1%
Mean absolute error	0.17
Root mean squared error	0.23
Total Number of Instances	4096

		Actual Class			
		Normal pump	Bearing	Screw	Gear
Predicted Class	Normal pump	1023	39	43	23
	Bearing	2	985	2	6
	Screw	0	0	978	0
	Gear	0	1	2	996

Fig.7. - Confusion matrix using hyperparameter optimization

True Positive Rate (TP Rate) / Sensitivity / Recall is the proportion of actual positive cases that were correctly identified by the model. A higher TP rate means the model is good at detecting positive instances (e.g., correctly identifying a fault when it is actually present).

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

False Positive Rate (FP Rate) represents proportion of negative cases that were incorrectly classified as positive. A lower FP rate is desirable as it indicates fewer false alarms.

$$FP\ Rate = \frac{False\ Positives}{True\ Negatives + False\ Positives}$$

Precision represents the proportion of predicted positive cases that were actually correct. Precision focuses on the quality of positive predictions. High precision means fewer false positives.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

F-Measure (F1-Score) is the harmonic mean of Precision and Recall, providing a single metric that balances both. It is useful when you want a balance between precision and recall. A high F1-score means both false positives and false negatives are low.

$$\mathrm{F-Measure\ (F1-Score)} = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Matthews Correlation Coefficient (MCC) is a measure of the quality of binary classifications, considering all four confusion matrix categories: TP, TN, FP, and FN. It ranges from -1 to +1. A value of +1 indicates perfect prediction, 0 indicates random prediction, and -1 indicates total disagreement

$$\mathrm{MCC} = \frac{(TF \times TN) - (FP \times FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$

ROC Area (Area Under the ROC Curve – AUC-ROC) is the area under the Receiver Operating Characteristic (ROC) curve, which plots TP Rate against FP Rate at various thresholds. It ranges from 0 to 1.Higher values (closer to 1) indicate better model discrimination ability between classes.

RC Area (Area Under the Precision-Recall Curve – AUC-PR)is the area under the curve that plots Precision against Recall for different threshold values. It ranges from 0 to 1. It is especially useful when dealing with imbalanced datasets. A high PRC area indicates the model maintains high precision and recall across thresholds.

Table 4. Detailed accuracy by class using hyperparameter optimization

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.954	0.000	1.000	0.954	0.977	0.969	1.000	0.999	screw
0.998	0.034	0.907	0.998	0.950	0.935	0.998	0.995	normal
0.972	0.001	0.997	0.972	0.984	0.979	0.998	0.995	gear
0.961	0.003	0.990	0.961	0.975	0.967	0.999	0.998	bearing
0.971	0.010	0.973	0.971	0.972	0.963	0.999	0.997	Weighted Avg

4. Results discussion

The experimental results from this study clearly demonstrate the efficacy of various machine learning classifiers in identifying mechanical faults using vibration signals. The performance was evaluated across five common classifiers—Naïve Bayes, Logistic Regression, Decision Table, Random Tree, and Random Forest—using both multiclass classification (four fault conditions) and binary classification (normal vs. each fault type). Furthermore, Auto-WEKA was employed to optimize model hyperparameters, resulting in a significant boost in predictive accuracy.

Table 1 presents the classification accuracy for the multiclass fault diagnosis task using unoptimized models. Among the five classifiers; Naïve Bayes achieved the lowest accuracy (28%), primarily due to its strong assumption of feature independence, which is often violated in vibration data. Logistic Regression and Decision Table showed moderate improvement, with accuracies of 36% and 42%, respectively. Random Tree outperformed the earlier models with a 45% accuracy. Random Forest achieved the highest accuracy of 53%, confirming its strength in handling complex patterns through ensemble learning. However, even the best performance from unoptimized models remained suboptimal, likely due to default parameter settings and the complexity of distinguishing between similar vibration patterns in mechanical components.

In binary classification tasks (Table 2), where normal conditions were compared individually against looseness, bearing faults, and gear faults. All classifiers showed improved accuracy compared to the multiclass task. Random Forest again outperformed other models, achieving up to 81.7% accuracy when distinguishing normal from gear faults. The accuracy declined slightly for bearing faults (70.8%) and looseness faults (77.9%), indicating overlapping features between these fault types and healthy conditions. These results suggest that binary classification simplifies the decision boundaries, allowing models to focus on more distinct patterns between two classes. However, this approach is less practical in real-world settings, where simultaneous detection of multiple faults is required.

To address the limitations of default settings, Auto-WEKA was employed to perform algorithm selection and hyperparameter tuning. The optimized Random Forest model was configured with 130 trees and a maximum depth of 15, enabling it to capture deeper interactions without overfitting. As shown in Table 3, the optimized model achieved a classification accuracy of 97.1% on the full dataset (four-class classification), a dramatic improvement from the initial 53%. The estimated error rate was reduced to 0.85%, signifying highly reliable classification and

further confirming the robustness of the optimized model.

Figure 7 illustrates the confusion matrix, showing that most samples were correctly classified across all categories. The model showed near-perfect classification of screw looseness and gear faults, with almost no misclassifications. Some misclassification occurred between normal and bearing faults, which may be attributed to subtle vibration differences in early bearing degradation. The class-wise evaluation metrics in Table.3 shows that the Precision and Recall values exceeded 0.95 across all fault types. The F1-score, which balances precision and recall, was above 0.97 for most classes. The Matthews Correlation Coefficient (MCC) values, a balanced measure even for imbalanced data, were all above 0.96, indicating strong predictive reliability. Both the ROC Area and PRC Area were close to 1.0, reflecting excellent separability between classes. These results validate the effectiveness of using optimized Random Forest models for real-time fault detection with limited data. The model not only generalizes well but also maintains high sensitivity and specificity across all fault categories.

However, limitations remain such as the system was tested under controlled laboratory conditions; field environments may introduce noise and variability not captured here.

Future work should explore transfer learning or domain adaptation to apply the model across different machines or operational scenarios. Incorporating time-domain features or additional sensor modalities (e.g., acoustic or thermal signals) could further enhance diagnostic robustness.

Conclusion

This study demonstrated that machine learning algorithms, particularly the Random Forest model, are highly effective in diagnosing mechanical faults using vibration signals. The experimental setup enabled the generation of a compact yet diverse dataset encompassing four operational conditions. Initial tests across multiple classifiers showed moderate performance, with Random Forest achieving the highest accuracy at 53% before optimization. By applying Auto-WEKA for hyperparameter tuning and algorithm selection, the optimized Random Forest model achieved a significantly improved classification accuracy of 97.1% and a low error rate of 0.85%. This confirms that hyperparameter optimization can greatly enhance model performance, even when using limited datasets. The outcome supports the integration of intelligent diagnostic systems in predictive maintenance applications, improving system reliability while minimizing downtime and inspection costs. These findings highlight the effectiveness of optimized ensemble learning techniques for early fault detection in mechanical systems with limited data.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This work was funded by Sultan Qaboos University-Qatar University joint fund Grant No. CL/SQU\QU/ENG/24/01.

References

[1] Kankar P., Sharma S. C., Harsha S. P. Fault diagnosis of ball bearings using machine learning methods //Expert Systems with Applications, 2011, vol. 38, no. 3, pp. 1876–1886
[2] Khan S., Yairi T. A review on the application of deep learning in system health management //Mechanical Systems and Signal Processing, 2018, vol. 107, pp. 241–265
[3] Jia F., Lei Y., Lin J., Zhou X., Lu N. Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data //Mechanical Systems and Signal Processing, 2016, vol. 72-73, pp. 303–315
[4] Lei Y., He Z., Zi Y. A new approach to intelligent fault diagnosis of rotating machinery //Expert Systems with Applications, vol. 2008, 35, no. 4, pp. 1593–1600
[5] Bouvrie J. Notes on convolutional neural networks //Tech. Rep., MIT, 2006.
[6] Chen Z., Li C., Sanchez R. V. Gearbox fault identification and classification with convolutional neural networks //Shock and Vibration, 2015.
[7] Gude A., Pawar S., Alhat S., Mishra S. Vibration analysis for engine fault detection //Journal of Robotics and Control (JRC), 20212(3), 205-208.
[8] Berenji A., Taghiyarrenani Z., Bastami A. R. Fault identification with limited labeled data //*Journal of Vibration and Control*. 2023, 30(7-8),1502-1510
[9] Li C., Gilad A., Glavic B., Miao Z., Roy, S. Refining labeling functions with limited labeled data // arXiv:2505.23470
[10] Gilanifar, M., Wang, H., Cordova, J., Ozguven, E. E., Strasser, T. I., Arghandeh R. Fault classification in power distribution systems based on limited labeled data using multi-task latent structure learning //Sustainable Cities and Society, 2021, 73, 1–10.
[11] Jin Y., Hou L., Du M., Chen Y. A limited labeled data bearing fault diagnosis method based on self-supervised learning. arXiv. https://arxiv.org/abs/2207.10432, 2022
[12] Zhao R., Yan R., Chen Z., Mao K., Wang P., Gao R. X. Iterative deep graph learning and embedding for industrial fault diagnosis with limited labeled data //IEEE Transactions on Industrial Informatics, 2021, 17(6), 4090–4099.
[13] Thornton C., Hutter F., Hoos H. H., Leyton-Brown K. Auto-WEKA: Combined selection and hyperparameter optimization of classification algorithms // Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2013, pp. 847–855.

Information of the authors

Zarog Musaab, PhD, assistant professor, Department of Mechanical and Industrial Engineering, College of Engineering, Sultan Qaboos University
e-mail: musaabh@squ.edu.om

Al-Adawi Sulaiman, BEng, graduate student, Department of Mechanical and Industrial Engineering, College of Engineering, Sultan Qaboos University
e-mail: s131265@student.squ.edu.om

Al-Nabhani Mohammed, BEng, graduate student, Department of Mechanical and Industrial Engineering, College of Engineering, Sultan Qaboos University
e-mail: s131265@student.squ.edu.om

Al-Aufi Isehaq, BEng, graduate student, Department of Mechanical and Industrial Engineering, College of Engineering, Sultan Qaboos University
e- mail: s131265@student.squ.edu.om