

Mohammad HAJHOSSEINI ¹, Ali DORKHAH ²,
Mohammad HOSSEINI ², Mohammad Sadegh JAFARI ³

Vibration spectrogram-based ResNet–LSTM for bearing fault detection and severity estimation

Received 26 May 2025, Revised 7 July 2026, Accepted 13 August 2026, Published online

Keywords: rolling element bearings, fault detection, severity classification, spectrogram representation, CNN-LSTM

Accurate and reliable fault diagnosis of rolling element bearings is essential for ensuring the safety and efficiency of rotating machinery. Traditional methods often rely on handcrafted features, which are limited in their ability to capture the complex and nonstationary nature of vibration signals. In this study, a deep learning framework that combines spectrogram representations of vibration signals with a hybrid ResNet–LSTM architecture is proposed to detect fault type and classify severity level, simultaneously. Time-domain vibration signals are first converted into spectrograms to preserve both temporal and frequency-domain characteristics. These spectrograms are then processed as sequences, enabling the ResNet component to extract discriminative spatial features while the LSTM captures temporal dependencies across frames. The proposed model is trained and validated under different sequence lengths, strides, and LSTM unit sizes. Evaluation of the model on test data demonstrates that the near-perfect accuracy (> 0.99) in fault detection and high accuracy (0.91–0.99) in severity level classification is achieved using the proposed framework. Evaluation of the model on the industrial case study further shows that the proposed model not only achieves high accuracy on controlled datasets but also provides reliable predictions under realistic operating conditions.

✉ Mohammad HAJHOSSEINI, e-mail: m.hajhosseini@vru.ac.ir

¹Department of Mechanical Engineering, Vali-e-Asr University of Rafsanjan, Rafsanjan, Iran

²Flotation and Concentration Unit of Sarcheshmeh Copper Complex, Rafsanjan, Iran

³Head of Technical & Engineering Research Department, Sarcheshmeh Copper Complex, Rafsanjan, Iran



© 2026, The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution (CC-BY 4.0, <https://creativecommons.org/licenses/by/4.0/>), which permits use, distribution, and reproduction in any medium, provided that the author and source are cited.

1. Introduction

Bearings are among the most critical components of rotating machinery, ensuring smooth motion transfer and load support across aerospace, energy, transportation, and manufacturing industries. Their continuous operation under harsh and variable conditions makes them particularly susceptible to localized defects such as spalls and cracks. Even minor bearing faults can propagate rapidly, leading to catastrophic machine failures, costly downtime, and safety risks. Consequently, accurate bearing fault detection and severity estimation have become central to predictive maintenance strategies [1].

Several emerging sensing technologies have demonstrated significant potential for bearing fault detection [2]. Alternative sensing modalities such as MEMS [3], piezoelectric-based systems [4], thermal imaging [5], and acoustic emission [6] have expanded the toolbox for bearing diagnostics, especially where vibration sensors are impractical.

While alternative sensing techniques have shown promise, vibration-based methods remain the foundational approach for bearing fault detection [7]. For example, McFadden [8] demonstrated early gear fault detection using amplitude and phase demodulation. Huang et al. [9] proposed empirical mode decomposition (EMD) for analyzing nonlinear and nonstationary vibration signals. These classical approaches often relied on handcrafted features and conventional classifiers such as support vector machines [10]. Although useful, they were highly sensitive to noise, sensor placement, and operating variability, limiting their deployment in real-world conditions.

Deep learning approaches have transformed bearing fault diagnosis by enabling automatic feature extraction from raw vibration or time-frequency signals. Zhang et al. [11] proposed an STFT-CNN method of converting vibration signals into time-frequency images and achieved nearly 100% fault classification accuracy on both CWRU and MFPT datasets. Li et al. [12] proposed a CNN with global residual attention, demonstrating enhanced classification accuracy across multiple operating conditions by integrating attention mechanisms with convolutional feature extraction. Han et al. [13] introduced a CNN-LSTM-GRU hybrid model that captured both spatial features and temporal dependencies in vibration sequences, achieving robust fault detection even under variable load conditions.

Multimodal data fusion has further enhanced diagnostic reliability. Alam et al. [14] combined vibration and temperature signals, showing that multimodal analysis improved fault classification performance in industrial testbeds. Chen et al. [15] implemented a multimodal self-supervised learning framework for cross-domain bearing fault diagnosis, demonstrating strong generalization across different machines and operational regimes. Fu et al. [16] proposed an improved denoising and feature extraction method for bearing signals, which enhanced model robustness against noisy measurements.

Attention-augmented and recurrent architectures have been extensively studied. DengFeng et al. [17] combined multi-scale CNNs with BiLSTM and attention mechanisms, showing superior fault diagnosis accuracy under varying speed and load conditions. Mirzaeibonehkhater et al. [18] developed a Transformer-based CNN-LSTM model that captured long-range dependencies, achieving state-of-the-art performance in benchmark datasets. Liu et al. [19] applied LSTM-based statistical process analysis to predict bearing failures, highlighting the importance of temporal modeling in fault prognosis.

Robust deployment in real-world environments has been a focus. Xu et al. [20] designed an anti-noise, easily deployable detection method that maintains high accuracy under heavy signal interference. Yang et al. [21] developed a hybrid CNN-BiLSTM-MHSA model that demonstrated precise fault diagnosis and improved feature representation across multiple datasets. Domingues et al. [22] leveraged self-supervised learning to generate effective feature embeddings, reducing dependence on labeled data and enhancing cross-domain adaptability. Wang et al. [23] introduced a deep learning framework for rolling bearing fault diagnosis that improved classification performance using lightweight architectures optimized for online monitoring.

Transformer-based and self-supervised learning models have further advanced the field. Yang et al. [24] employed a lightweight transformer for efficient fault detection, achieving competitive accuracy with reduced computational cost. Zhao et al. [25] provided a comprehensive review of deep learning-based fault diagnosis, highlighting recent advances and remaining challenges. Ahmed et al. [26] proposed a convolutional-transformer model with long-range attention that effectively captured temporal-spatial features for early fault detection. Luo et al. [27] applied masked self-supervised learning for unsupervised bearing fault diagnosis, showing strong performance without labeled data. Chen et al. [28] introduced an efficient cross-space multiscale CNN-transformer parallelism approach that demonstrates precise fault localization and severity estimation. Siddique et al. [29] combined continuous wavelet transform with attention-enhanced spatiotemporal feature extraction, achieving high accuracy for complex vibration signals.

Despite these advances, several challenges remain. Many deep learning models are still benchmarked primarily on public datasets (e.g., CWRU), which do not fully reflect the variability of real industrial environments. Although CNNs are highly effective in extracting spatial features, they are less capable of modeling long-term temporal dependencies, thereby motivating the use of hybrid architectures. Furthermore, severity estimation has received considerably less attention than fault classification, despite its importance for predictive maintenance and maintenance decision-making.

Although considerable progress has been achieved in intelligent bearing fault diagnosis, several research gaps remain. Existing studies mainly focus on fault classification, while relatively few have addressed the simultaneous prediction of fault type and severity within a unified deep learning framework. Moreover, many

spectrogram-based methods process individual images independently without fully exploiting the temporal relationships between consecutive vibration patterns. In addition, validation is often limited to laboratory benchmark datasets, whereas comprehensive evaluation using real industrial vibration data remains scarce. These limitations highlight the need for a more robust and practical framework for intelligent bearing condition monitoring.

Motivated by these research gaps, this study proposes a vibration spectrogram-based ResNet–LSTM framework for the simultaneous detection and severity estimation of bearing faults. The proposed approach converts raw vibration signals into sequences of spectrogram images, enabling the ResNet backbone to extract discriminative spatial features while the LSTM captures temporal dependencies between consecutive spectrograms. In addition, a systematic sensitivity analysis is conducted to assess the effects of sequence length, stride, and the number of LSTM units on diagnostic performance. Finally, the proposed framework is validated using both a benchmark bearing dataset and two industrial case studies involving rougher bearings and housing pump bearings measurements collected from the Sarcheshmeh Copper Complex, demonstrating its effectiveness and practical applicability under realistic operating conditions.

2. Methodology

2.1. Dataset description

The University of Ottawa Rolling-element Dataset – Vibration and Acoustic Faults under Constant Load and Speed conditions (UORED-VAFCLS) is employed as the primary dataset for model development and validation [30]. The dataset contains vibration measurements acquired from rolling-element bearings operating under constant load and rotational speed conditions. It includes five bearing health categories: healthy bearings, inner race faults, outer race faults, ball faults, and cage faults. Multiple vibration measurements are provided for each fault category, allowing the development and evaluation of machine learning models under different bearing conditions. Each sample is provided in raw time-domain form with a sampling frequency of 42 kHz.

In this study, only the raw time-domain vibration signals are used, rather than the pre-computed spectrograms available in the dataset. This decision is motivated by the need to maintain a consistent transformation pipeline between the benchmark dataset and the plant-measured signals used later for generalization testing.

Besides the fault-type labels, the benchmark dataset also includes three predefined severity classes that are used as the target labels for the severity estimation task. In this study, severity level 0 represents the healthy bearing condition, whereas severity levels 1 and 2 correspond to moderate and severe fault conditions, respectively. These severity labels are directly employed during the supervised training of the proposed ResNet–LSTM framework. During the industrial case study, vi-

bration signals acquired by the Condition Monitoring (CM) unit are processed using the same preprocessing procedure, and the trained model estimates the corresponding severity level based on the severity classification criteria learned from the benchmark dataset.

2.2. Transformation of raw signals into spectrograms

Although spectrogram images are available in the benchmark dataset, new spectrograms are generated from every raw vibration signal to ensure a consistent preprocessing procedure for both the benchmark dataset and the plant-measured signals used in the industrial validation. Each signal is first preprocessed by removing its DC component and applying RMS-based scaling using the root-mean-square value calculated across the dataset. This preprocessing ensures comparable signal amplitudes and improves the consistency of model training.

After preprocessing, each vibration signal is divided into 400 equal segments. An STFT-based spectrogram is computed for every segment using a window length of 256 samples with 50% overlap between adjacent windows. The spectrogram magnitudes are then converted to a logarithmic scale and represented as RGB images using the plasma colormap. Finally, each spectrogram is resized to 224×224 pixels, corresponding to the input size required by the ResNet50V2 backbone.

The complete preprocessing workflow is illustrated in Fig. 1, which summarizes the successive processing steps from raw vibration signals to the final spectrogram images used as inputs to the proposed ResNet–LSTM framework. The entire procedure is implemented automatically using a Python script. Representative spectrograms generated using this workflow are presented in Figs. 2 and 3 for healthy and cage-fault conditions, respectively.

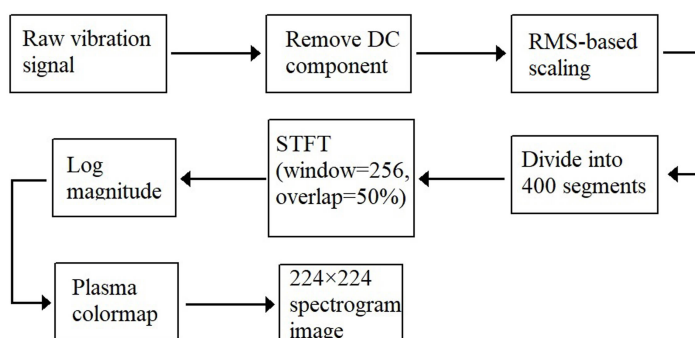


Fig. 1. Workflow for transforming raw vibration signals into spectrogram images used as inputs to the proposed ResNet–LSTM framework

Bearing defects generate characteristic vibration frequencies that are determined by the bearing geometry and shaft rotational speed. These include the ball pass frequency of the outer race (BPFO), ball pass frequency of the inner race

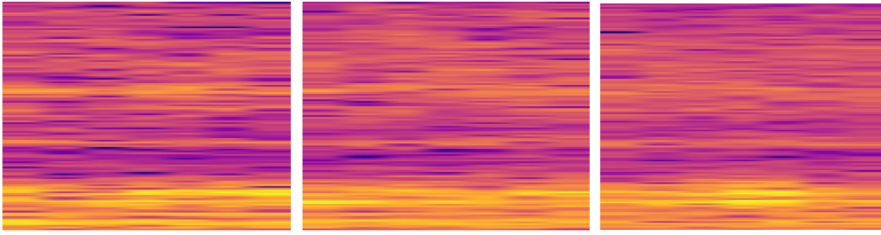


Fig. 2. Example spectrograms for healthy condition

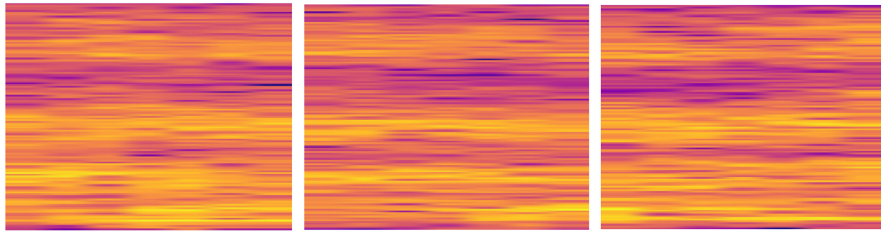


Fig. 3. Example spectrograms for cage-fault condition

(BPFI), ball spin frequency (BSF), and fundamental train frequency (FTF), which are commonly used in conventional bearing fault diagnosis. Rather than explicitly calculating these characteristic frequencies, the proposed framework learns the corresponding fault-related time-frequency patterns directly from the STFT-based spectrograms. In the spectrogram, these characteristic frequency components appear as distinctive time-frequency energy patterns, where the color intensity represents the magnitude (energy) of the corresponding frequency component at each time instant. The ResNet50V2 backbone automatically extracts discriminative spatial features from these patterns, while the LSTM layer models their temporal evolution across consecutive spectrogram images. Consequently, the proposed ResNet–LSTM framework can automatically distinguish different bearing fault types and estimate their severity without requiring explicit calculation of characteristic frequencies or handcrafted frequency-domain features.

2.3. Sequence preparation

For each signal, the spectrogram images are grouped into overlapping sequences. Each sequence consists of a fixed number of consecutive spectrograms, referred to as the sequence length. The window used to extract the next sequence is shifted forward by a specified number of images, referred to as the stride. For example, Fig. 4 illustrates the case where the sequence length is set to five and the stride to two. The first sequence containing spectrograms one through five is highlighted with a red rectangle. The second sequence begins two images later is highlighted with a black rectangle. This sliding-window procedure creates overlap-

ping segments that preserve local temporal continuity while providing sufficient coverage of the entire signal. A smaller stride increases overlap and thus improves temporal resolution, whereas a larger stride reduces redundancy but may overlook finer variations in the signal.

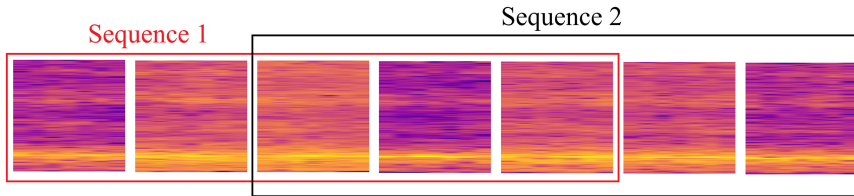


Fig. 4. Construction of spectrogram sequences with sequence length = 5 and stride = 2

2.4. Model architecture

The proposed diagnostic framework combines a convolutional neural network (CNN) with long short-term memory (LSTM) units in a two-stage architecture. A sequence of spectrogram images is first processed by a TimeDistributed CNN, where a pre-trained ResNet50V2 backbone serves as the feature extractor. The convolutional layers of ResNet50V2 extract discriminative time–frequency features from each individual spectrogram frame, and the resulting feature maps are compressed using a global average pooling layer to produce compact frame-level feature representations.

These frame-level representations are subsequently processed by an LSTM layer, which models temporal dependencies across consecutive spectrogram frames and captures the evolution of fault-related characteristics over time. To improve generalization and reduce overfitting, a dropout layer is applied after the LSTM output. The resulting shared feature representation is then fed into two parallel softmax classification heads: one for fault type classification and the other for fault severity estimation.

The proposed ResNet–LSTM framework adopts a shared-feature multi-task learning architecture, in which the ResNet50V2 backbone, the LSTM layer, and the dropout layer are shared between the fault type classification and severity estimation tasks. The resulting shared feature representation is subsequently fed into two independent Dense layers with softmax activation to generate the fault type and severity predictions. This hard parameter-sharing strategy enables both tasks to learn a common discriminative feature representation while maintaining separate task-specific output layers. During supervised training, the framework adopts a multi-task learning objective in which sparse categorical cross-entropy loss is computed independently for each output branch. The network parameters are optimized by minimizing the equally weighted sum of the two losses, and no task-specific loss weighting is introduced. The overall architecture of the proposed framework is illustrated in Fig. 5.

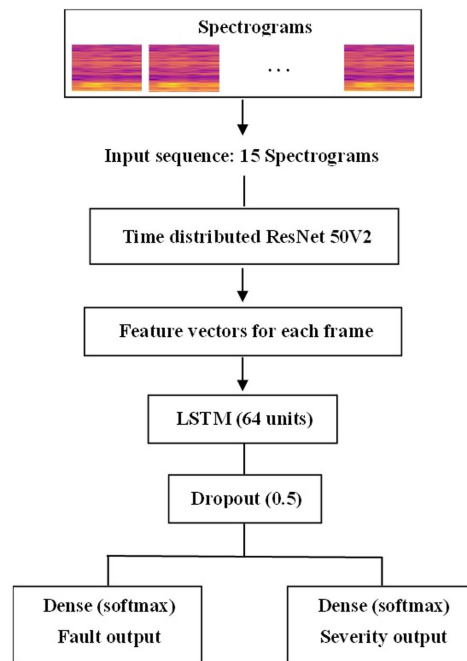


Fig. 5. Architecture of the proposed diagnostic model

3. Results and discussion

3.1. Hyperparameter sensitivity analysis

In order to examine the influence of the design parameters on the model performance, a hyperparameter sensitivity analysis is conducted. The parameters under investigation include the sequence length, stride, and the number of LSTM units. The sequence length determines how many consecutive spectrogram images are grouped into a single sequence, while the stride controls the degree of overlap between successive sequences. The number of LSTM units determines the model's capacity to capture temporal dependencies.

For each hyperparameter configuration, the complete dataset is randomly divided into training, validation, and testing subsets using a stratified sampling strategy to preserve the class distribution of the fault categories. Approximately 64% of the total data is used for training, 21% for validation, and 15% for independent testing. The data partition is generated using a fixed random seed (random state = 42) and remains unchanged throughout the hyperparameter sensitivity analysis to ensure a fair comparison among all configurations. The CNN–LSTM model is trained using the training set, validated using the validation set, and finally evaluated on the independent test set. Accuracies are measured separately for fault type classification and severity estimation to ensure that the reported results reflect the

model's generalization capability rather than its training performance. Results are listed in Table 1. This table illustrates that, for each combination of LSTM units and sequence length, the severity estimation accuracy increases as the stride decreases. In other words, a small stride proves essential for precise severity estimation, despite its higher computational cost. Figs. 6 and 7 focus exclusively on the stride = 1 cases, illustrating how the sequence length affects accuracy for models with 64 and 128 LSTM units. These figures clearly confirm that fault accuracy is nearly

Table 1. Hyperparameter Sensitivity Analysis

LSTM units	Sequence length	Stride	Fault accuracy (%)	Severity accuracy (%)
64	5	1	0.9983	0.9677
		3	0.9882	0.9066
	10	1	0.9997	0.9935
		3	0.9966	0.9644
		5	0.9916	0.9395
	15	1	1.0000	0.9957
		3	0.9991	0.9509
		5	0.9929	0.9416
128	5	1	0.9986	0.9854
		3	0.9907	0.9175
	10	1	1.0000	0.9955
		3	0.9941	0.9695
		5	0.9958	0.9381
	15	1	0.9997	0.9686
		3	0.9974	0.9664
		5	0.9972	0.9601

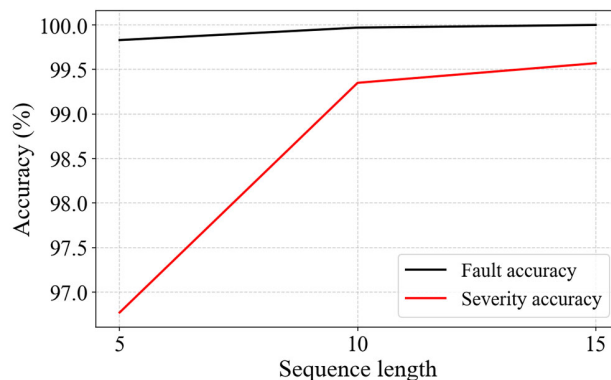


Fig. 6. Effects of sequence length on accuracies for 64 LSTM units and stride = 1

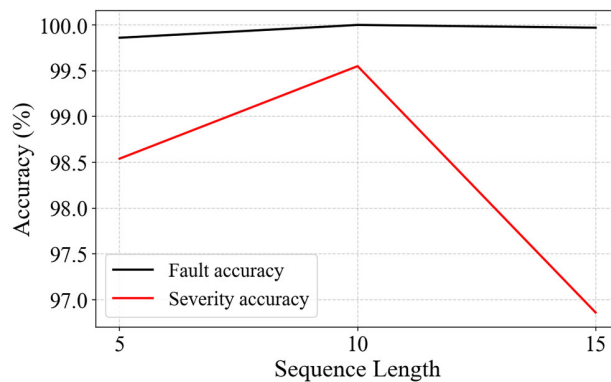


Fig. 7. Effects of sequence length on accuracies for 128 LSTM units and stride = 1

saturated across all sequence lengths, while severity accuracy benefits substantially from longer sequences before leveling off.

The number of LSTM units affects performance only marginally. Increasing units from 64 to 128 provides modest gains for shorter sequences, but at longer sequence lengths the difference is negligible or even detrimental, indicating diminishing returns.

Overall, the sensitivity analysis shows that sequence length is the most critical parameter, followed by stride, with LSTM units playing a secondary role. Based on test-set results, the optimal configuration is identified as sequence length = 15, stride = 1, and 64 LSTM units, which achieves 100% fault accuracy and 99.6% severity accuracy. The final architecture and training hyperparameters of this optimal model are summarized in Table 2. This model is then saved and used in

Table 2. Final architecture and training hyperparameters of the proposed ResNet–LSTM model

Hyperparameter	Final value
Input image size	224 × 224 pixels
CNN backbone	ResNet50V2 (ImageNet pretrained)
Sequence length	15
Sequence stride	1
LSTM units	64
Dropout rate	0.5
Optimizer	Adam
Initial learning rate	0.0001
Batch size	32
Maximum training epochs	200
Early stopping patience	25
Loss function	Sparse categorical cross-entropy

the subsequent confusion matrix analysis (Section 3.2) and industrial validation (Section 4).

3.2. Confusion matrix of the best model

The saved best model with optimal configuration of hyperparameters is evaluated on the test set, and its performance is summarized using both training curves and confusion matrices. Fig. 8 illustrates the evolution of model accuracy during training. In this figure, part (a) shows training and validation accuracies for fault classification across epochs, and part (b) shows the corresponding curves for severity estimation. Fig. 9 presents the total training and validation losses over epochs, confirming stable convergence without signs of overfitting.

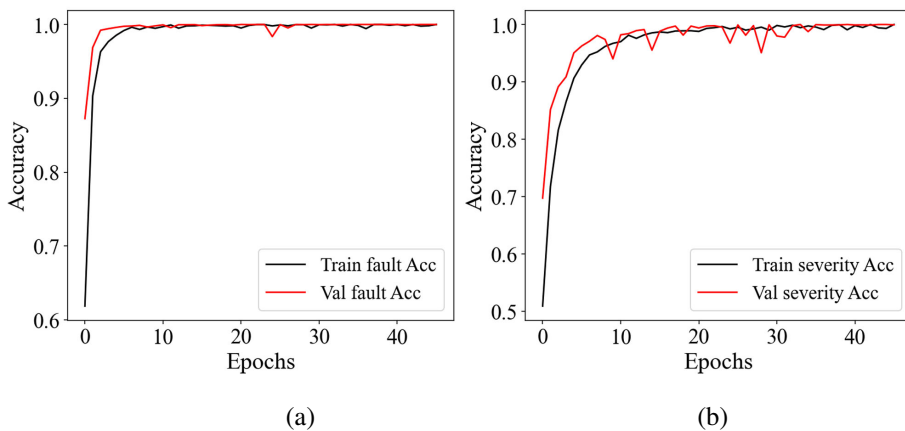


Fig. 8. Train and validation accuracies for: (a) fault classification, (b) severity estimation

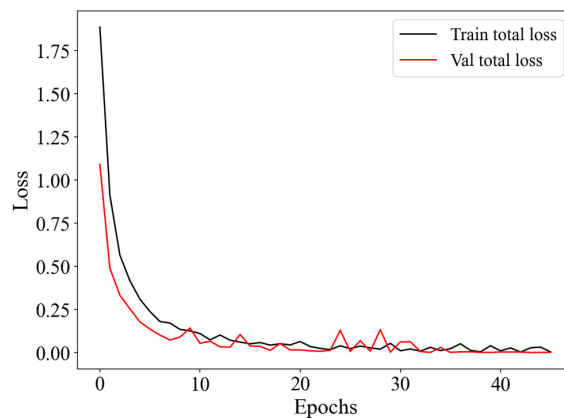


Fig. 9. Train and validation losses over the epochs for the best model

To further validate the generalization ability, confusion matrices are provided in Figs. 10 and 11. The fault classification matrix demonstrates perfect separation across all conditions, confirming the robustness of spectrogram-based sequence modeling. The severity classification matrix also shows high accuracy, though with some minor confusion between adjacent severity levels, reflecting the inherent challenge of fine-grained estimation.

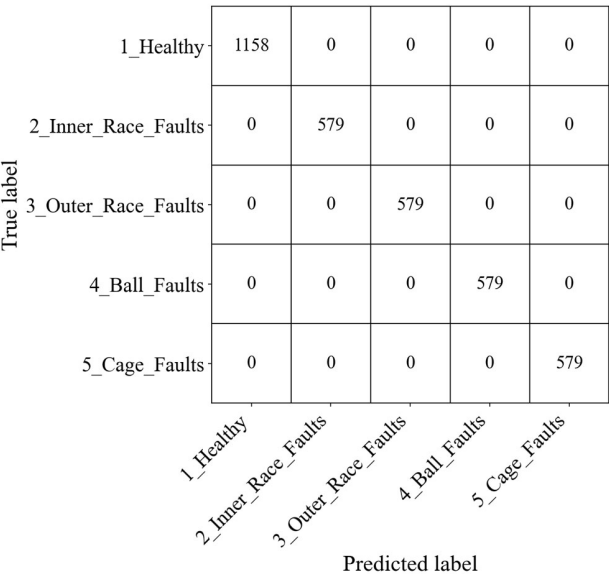


Fig. 10. Confusion matrix for fault classification

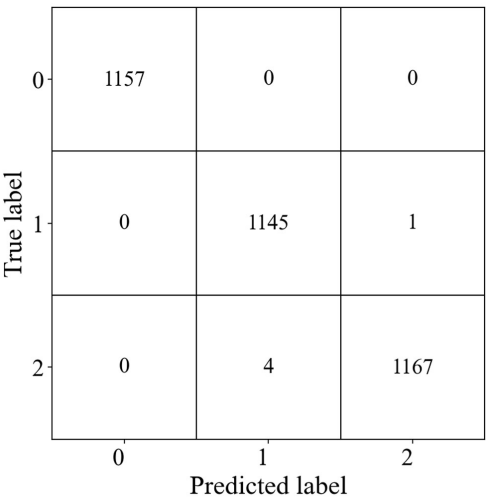


Fig. 11. Confusion matrix for severity estimation

3.3. Comparison with state-of-the-art methods

To further evaluate the proposed framework, its characteristics are compared with several recently published deep learning-based bearing fault diagnosis methods. The selected studies represent recent CNN-, LSTM-, and Transformer-based approaches that have demonstrated excellent diagnostic performance for rolling bearing fault diagnosis. Although direct numerical comparison of diagnostic accuracy is difficult because different benchmark datasets, fault categories, and experimental protocols were employed, comparing the capabilities of these methods provides a meaningful assessment of the practical advantages of the proposed framework.

Table 3 summarizes the main characteristics of representative state-of-the-art methods. Most existing studies focus primarily on bearing fault classification using laboratory datasets, while severity estimation and validation using real industrial vibration measurements are generally not considered. In contrast, the proposed ResNet–LSTM framework adopts a shared-feature multi-task learning architecture that simultaneously performs fault type classification and severity estimation within a unified model. Furthermore, unlike previous studies, the proposed framework is validated using vibration measurements collected from an operating industrial plant, demonstrating its applicability under realistic operating conditions.

Table 3. Comparison of the proposed method with recent state-of-the-art deep learning approaches for bearing fault diagnosis

Reference	Deep learning model	Simultaneous fault classification and severity estimation	Severity estimation	Industrial validation
Ahmed and Nandi [26]	CNN–Transformer	No	No	No
Han et al. [13]	CNN–LSTM–GRU	No	No	No
Yang et al. [21]	CNN–BiLSTM–MHSA	No	No	No
Luo and Liu [27]	Swin Transformer	No	No	No
Siddique et al. [29]	Attention-enhanced deep learning with CWT	No	No	No
This study	Shared-feature ResNet–LSTM	Yes	Yes	Yes

4. Case study: Validation with industrial plant data

In order to assess the generalization capability of the proposed ResNet–LSTM framework beyond the University of Ottawa benchmark dataset, industrial vibration measurements are collected from two rougher bearings and two housing pump

bearings at the Sarcheshmeh Copper Complex. The industrial validation dataset includes one healthy and one faulty rougher bearing, together with one healthy and one faulty housing pump bearing. The rougher bearings are equipped with SKF 6317 M deep-groove ball bearings, whereas the housing pump bearings are equipped with Timken 6317 deep-groove ball bearings. Although these industrial bearings differ in manufacturer and geometry from the NSK 6203ZZ and FAFNIR 203KD deep-groove ball bearings used in the University of Ottawa benchmark dataset, they belong to the same bearing family and exhibit the same fundamental fault mechanisms, including inner race, outer race, ball, and cage defects. Therefore, the industrial case studies provide an independent evaluation of the proposed ResNet–LSTM framework on different deep-groove ball bearing configurations under realistic operating conditions.

For each bearing condition, seven independent vibration measurements are collected, resulting in a total of 28 vibration measurements for industrial validation. As illustrated in Figs. 12 and 15, vibration signals are acquired using an SPM Instrument Leonova Diamond vibration analyzer under normal production operating conditions. The time-domain vibration signals are recorded over a duration of 10 s using the long-wave acquisition mode, and the sampling frequency is adjusted so that each signal contains the same number of samples as those in the benchmark dataset.



Fig. 12. Experimental setup for vibration data acquisition from the rougher bearing

The acquired time-domain signals are converted into spectrograms following the same procedure described in Section 2.2. Sequences are then generated using the optimal configuration identified in Section 3.1 (sequence length = 15, stride = 1, and

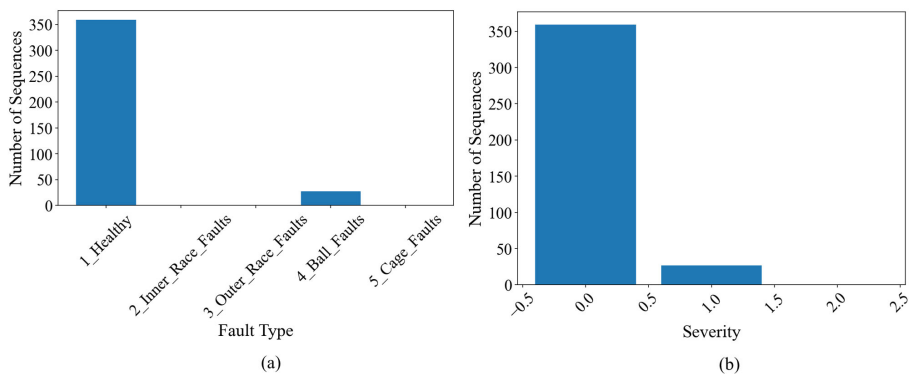


Fig. 13. Prediction distribution of spectrogram sequences for the healthy rougher bearing using the trained ResNet–LSTM model: (a) fault type, (b) severity level

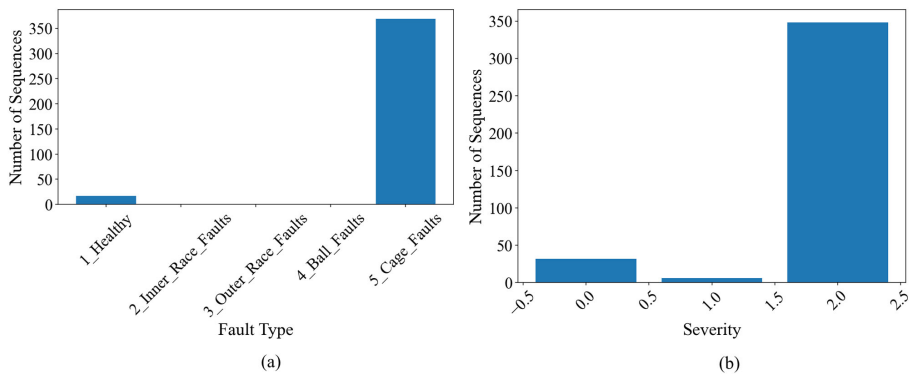


Fig. 14. Prediction distribution of spectrogram sequences for the faulty rougher bearing using the trained ResNet–LSTM model: (a) fault type, (b) severity level

64 LSTM units). For each rougher bearing, the number of sequences corresponding to healthy, faulty, and severity conditions are summarized in Figs. 13 and 14. Similarly, the prediction distributions obtained for the housing pump bearing are presented in Figs. 16 and 17.

For industrial validation, each spectrogram sequence is evaluated using the same trained ResNet–LSTM model developed from the University of Ottawa benchmark dataset. No additional training, fine-tuning, or parameter adjustment is performed using the industrial measurements. Instead, the industrial vibration signals undergo exactly the same preprocessing, spectrogram generation, sequence construction, and inference procedure described in Sections 2.2–2.4.

The prediction distributions for the rougher bearing are presented in Figs. 13 and 14. Under the healthy operating condition (Fig. 13), the proposed ResNet–LSTM framework correctly classifies nearly all spectrogram sequences as healthy and assigns the majority of sequences to severity level 0. Under the faulty operating condition (Fig. 14), most sequences are classified as cage faults and assigned to



Fig. 15. Experimental setup for vibration data acquisition from the housing pump bearing

severity level 2, with only a small number of misclassified sequences. These results demonstrate reliable fault identification and severity estimation for the rougher bearing under industrial operating conditions.

The prediction distributions for the housing pump bearing are presented in Figs. 16 and 17. For the healthy operating condition (Fig. 16), almost all spectrogram sequences are correctly classified as healthy, while the corresponding severity predictions are mainly distributed between severity levels 0 and 1. For the faulty operating condition (Fig. 17), the proposed framework correctly identifies the majority of sequences as ball faults and assigns almost all of them to severity level 2. These results further confirm the robustness of the proposed framework under different industrial operating conditions.

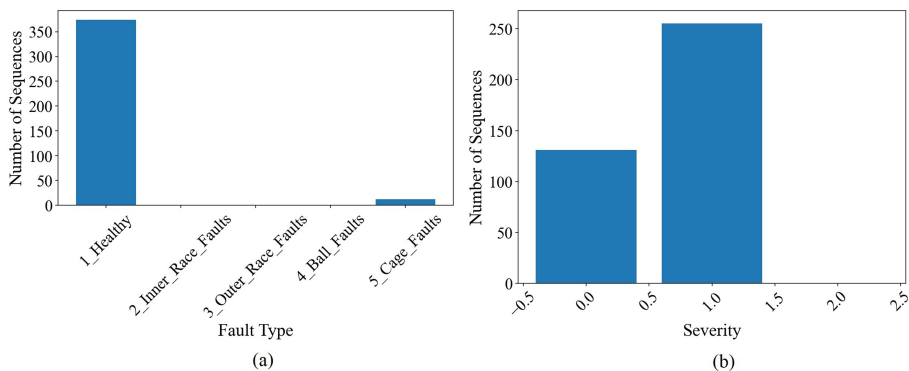


Fig. 16. Prediction distribution of spectrogram sequences for the healthy housing pump bearing using the trained ResNet–LSTM model: (a) fault type, (b) severity level

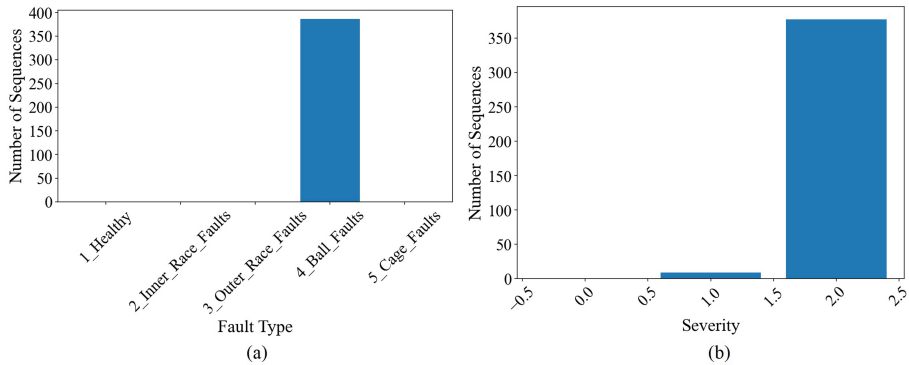


Fig. 17. Prediction distribution of spectrogram sequences for the faulty housing pump bearing using the trained ResNet–LSTM model: (a) fault type, (b) severity level

To further verify the industrial validation results, the model predictions are compared with the diagnoses provided by the Condition Monitoring (CM) experts and the observations obtained during machine overhaul. For the rougher bearing, the healthy operating condition identified by the CM experts is correctly classified by the proposed framework as healthy with severity level 0. For the faulty rougher bearing, the CM experts identify a cage fault, which is subsequently confirmed during machine overhaul as a medium-to-severe cage defect. Consistent with these findings, the proposed framework correctly predicts the cage-fault category and assigns the majority of spectrogram sequences to severity level 2. Similarly, for the housing pump bearing, the healthy operating condition identified by the CM experts is correctly classified as healthy with severity level 0. For the faulty housing pump bearing, the CM experts identify a ball fault, which is confirmed during machine overhaul as a severe ball defect. The proposed framework likewise predicts the ball-fault category and assigns almost all spectrogram sequences to severity level 2. The strong agreement between the model predictions, the CM expert assessments, and the overhaul observations demonstrates the reliability and practical applicability of the proposed framework for industrial bearing fault diagnosis and severity estimation.

Overall, the evaluation using plant-measured vibration data demonstrates that the proposed ResNet–LSTM framework can reliably detect both bearing fault type and severity under real industrial operating conditions. Unlike benchmark laboratory datasets, industrial vibration signals are affected by measurement noise, structural vibrations, load fluctuations, and interactions with neighboring mechanical components, making fault diagnosis considerably more challenging. Despite these factors, the proposed framework consistently produces reliable predictions for both rougher bearings and housing pump bearings, demonstrating its strong generalization capability beyond the benchmark training dataset. The limited number of misclassified sequences can be attributed to the similarity of vibration signatures under transitional operating conditions and the presence of unavoidable background

mechanical noise. Nevertheless, the dominant predictions remained consistent with the actual machine condition, indicating that the spectrogram representation, together with the ResNet–LSTM architecture, effectively captures both spatial and temporal fault characteristics. These results demonstrate the practical applicability of the proposed framework for predictive maintenance and intelligent condition monitoring in real industrial environments.

5. Conclusions

In this study, a deep learning framework for simultaneous fault type diagnosis and severity level classification of rolling element bearings is presented. By transforming raw vibration signals into spectrogram representations and modeling them as sequences, the proposed ResNet–LSTM architecture effectively captures both spatial and temporal dependencies inherent in the data. A systematic sensitivity analysis of sequence length, stride, and the number of LSTM units reveals that longer sequences and smaller strides significantly enhance severity estimation, while fault detection remains robust across most configurations. Using the optimal configuration of sequence length = 15, stride = 1, and 64 LSTM units, the proposed framework achieved 100% fault classification accuracy and 99.6% severity estimation accuracy on the benchmark test dataset.

Beyond controlled datasets, the framework is further validated using industrial case studies at the Sarcheshmeh Copper Complex, where plant-measured vibration signals are collected from both rougher bearings and housing pump bearings under healthy and faulty operating conditions. The proposed model demonstrates strong generalization capability, reliably predicting both fault presence and severity under realistic industrial environments.

The main contributions of this work can be summarized as follows:

- Development of a hybrid ResNet–LSTM architecture capable of joint fault type and severity level prediction.
- Implementation of a preprocessing strategy that removes the DC component and applies RMS-based scaling across the dataset, ensuring amplitude comparability and improving model training effectiveness.
- Introduction of a spectrogram-based sequential representation of vibration signals, enhancing temporal–frequency information capture.
- Comprehensive hyperparameter sensitivity analysis to guide optimal model design for bearing fault diagnosis.
- Validation on industrial plant data acquired from rougher bearings and housing pump bearings, confirming robustness and applicability beyond laboratory datasets.

Although the proposed ResNet–LSTM framework demonstrates excellent performance on both the benchmark dataset and the industrial case studies, several limitations should be acknowledged. First, the proposed framework is developed

and validated using vibration signals collected under a limited range of operating conditions and machine types. Its performance under substantially different operating speeds, loading conditions, and bearing configurations requires further investigation. Second, the proposed framework relies exclusively on vibration measurements, whereas integrating additional sensing modalities, such as acoustic emission, motor current, or temperature signals, may further enhance diagnostic robustness. Finally, the computational complexity associated with deep learning models may limit real-time deployment on resource-constrained edge devices. Future work will therefore focus on extending the framework to other rotating machinery components, incorporating domain adaptation techniques to improve generalization under varying operating conditions, investigating multi-modal data fusion, and developing lightweight model architectures for real-time industrial deployment.

References

- [1] A. Lakikza, H. Cheghib, and N. Kahoul. Optimized variational mode decomposition for improved bearing fault diagnosis and performance evaluation. *Archive of Mechanical Engineering*, 71(4):467–495, 2024. doi: [10.24425/ame.2024.152615](https://doi.org/10.24425/ame.2024.152615).
- [2] M. Pandiyan and T.N. Babu. Systematic review on fault diagnosis on rolling-element bearing. *Journal of Vibration Engineering and Technologies*, 12(7):8249–8283, 2024. doi: [10.1007/s42417-024-01358-4](https://doi.org/10.1007/s42417-024-01358-4).
- [3] G. Sharma, T. Kaur, S.K. Mangal, N.K. Dhiman, and G.L. Jat. MEMS approach for rolling bearing fault diagnosis using vibration signal analysis. *Journal of Vibration Engineering and Technologies*, 13(1):10, 2025. doi: [10.1007/s42417-024-01730-4](https://doi.org/10.1007/s42417-024-01730-4).
- [4] K. Wang, H. Long, D. Song, and H. Shariar. Energy harvesting microelectromechanical system for condition monitoring based on piezoelectric transducer ring. *Micromachines*, 16(6):602, 2025. doi: [10.3390/mi16060602](https://doi.org/10.3390/mi16060602).
- [5] O. Janssens, R. Schulz, V. Slavkovikj, K. Stockman, M. Loccufier, R. Van de Walle, and S. Van Hoecke. Thermal image-based fault diagnosis for rotating machinery. *Infrared Physics and Technology*, 73:78–87, 2015. doi: [10.1016/j.infrared.2015.09.004](https://doi.org/10.1016/j.infrared.2015.09.004).
- [6] D. Hou, H. Qi, H. Luo, C. Wang, and J. Yang. Comparative study on the use of acoustic emission and vibration analyses for the bearing fault diagnosis of high-speed trains. *Structural Health Monitoring*, 21(4):1518–1540, 2022. doi: [10.1177/14759217211036025](https://doi.org/10.1177/14759217211036025).
- [7] P. Khaire and V. Phalle. A smart fault identification system for ball bearing using simulation-driven vibration analysis. *Archive of Mechanical Engineering*, 70(2):247–270, 2023. doi: [10.24425/ame.2023.145583](https://doi.org/10.24425/ame.2023.145583).
- [8] P.D. McFadden. Detecting fatigue cracks in gears by amplitude and phase demodulation of the meshing vibration. *ASME Journal of Vibration Acoustics*, 108(2):165–170, 1986. doi: [10.1115/1.3269317](https://doi.org/10.1115/1.3269317).
- [9] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N. Yen, C.C. Tung, and H.H. Liu. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 454(1971):903–995, 1998. doi: [10.1098/rspa.1998.0193](https://doi.org/10.1098/rspa.1998.0193).
- [10] A. Widodo and B.S. Yang. Support vector machine in machine condition monitoring and fault diagnosis. *Mechanical Systems and Signal Processing*, 21(6):2560–2574, 2007. doi: [10.1016/j.ymssp.2006.12.007](https://doi.org/10.1016/j.ymssp.2006.12.007).

- [11] Q. Zhang and L. Deng. 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(2):795–811, 2023. doi: [10.1007/s11668-023-01616-9](https://doi.org/10.1007/s11668-023-01616-9).
- [12] X. Li, Y. Chen, and Y. Liu. A novel convolutional neural network with global perception for bearing fault diagnosis. *Engineering Applications of Artificial Intelligence*, 143:109986, 2025. doi: [10.1016/j.engappai.2024.109986](https://doi.org/10.1016/j.engappai.2024.109986).
- [13] K. Han, W. Wang, and J. Guo. Research on a bearing fault diagnosis method based on a CNN-LSTM-GRU model. *Machines*, 12(12):927, 2024. doi: [10.3390/machines12120927](https://doi.org/10.3390/machines12120927).
- [14] T.E. Alam, M.M. Ahsan, and S. Raman. Multimodal bearing fault classification under variable conditions: A 1D CNN with transfer learning. *Machine Learning with Applications*, 100682, 2025. doi: [10.1016/j.mlwa.2025.100682](https://doi.org/10.1016/j.mlwa.2025.100682).
- [15] X. Chen, Y. Xue, M. Huang, and R. Yang. Multi-modal self-supervised learning for cross-domain one-shot bearing fault diagnosis. *IFAC-PapersOnLine*, 58(4):746–751, 2024. doi: [10.1016/j.ifacol.2024.08.215](https://doi.org/10.1016/j.ifacol.2024.08.215).
- [16] L. Fu, Z. Ma, Y. Zhang, S. Wang, and L. Zhang. An improved bearing fault diagnosis method based on variational mode decomposition and adaptive iterative filtering (VMD-AIF). *Journal of Mechanical Science and Technology*, 37(4):1601–1612, 2023. doi: [10.1007/s12206-023-0303-2](https://doi.org/10.1007/s12206-023-0303-2).
- [17] Z. Dengfeng, T. Chaoyang, F. Zhijun, Z. Yudong, H. Junjian, and H. Wenbin. Multi-scale convolutional neural network combining BiLSTM and attention mechanism for bearing fault diagnosis under multiple working conditions. *Scientific Reports*, 15(1):13035, 2025. doi: [10.1038/s41598-025-96137-w](https://doi.org/10.1038/s41598-025-96137-w).
- [18] M. Mirzaebonehkhater, M.A. Labbaf-Khaniki, and M. Manthouri. Transformer-based bearing fault detection using temporal decomposition attention mechanism. *arXiv preprint arXiv:2412.11245*, 2024. doi: [10.48550/arXiv.2412.11245](https://doi.org/10.48550/arXiv.2412.11245).
- [19] K. Liu, C. Pan, F. Lei, D. Hu, and H. Zuo. Fault prediction of bearings based on LSTM and statistical process analysis. *Reliability Engineering and System Safety*, 214:107646, 2021. doi: [10.1016/j.ress.2021.107646](https://doi.org/10.1016/j.ress.2021.107646).
- [20] Z. Xu, Z. Jia, Y. Wei, S. Zhang, Z. Jin, and W. Dong. A strong anti-noise and easily deployable bearing fault diagnosis model based on time-frequency dual-channel transformer. *Measurement*, 236:115054, 2024. doi: [10.1016/j.measurement.2024.115054](https://doi.org/10.1016/j.measurement.2024.115054).
- [21] Z. Yang, W. Li, F. Yuan, H. Zhi, M. Guo, B. Xin, and Z. Gao. Hybrid CNN-BiLSTM-MHSA model for accurate fault diagnosis of rotor motor bearings. *Mathematics*, 13(3):334, 2025. doi: [10.3390/math13030334](https://doi.org/10.3390/math13030334).
- [22] R.K. Domingues, J.A. Cordioli, and D. Silva. Development of a methodology for diagnosing faults in bearings operating under variable operating conditions based on self-supervised learning. *Proceedings of the Annual Conference of the PHM Society*, 16(1), 2024. doi: [10.36001/phm-conf.2024.v16i1.4183](https://doi.org/10.36001/phm-conf.2024.v16i1.4183).
- [23] Y. Wang, D. Li, L. Li, R. Sun, and S. Wang. A novel deep learning framework for rolling bearing fault diagnosis enhancement using VAE-augmented CNN model. *Heliyon*, 10(15), 2024. doi: [10.1016/j.heliyon.2024.e35407](https://doi.org/10.1016/j.heliyon.2024.e35407).
- [24] P. Yang, B. Zhang, and J. Zhao. Efficient fault diagnosis in rolling bearings lightweight hybrid model. *Scientific Reports*, 15(1):11514, 2025. doi: [10.1038/s41598-025-96285-z](https://doi.org/10.1038/s41598-025-96285-z).
- [25] J. Zhao, W. Wang, J. Huang, and X. Ma. A comprehensive review of deep learning-based fault diagnosis approaches for rolling bearings: Advancements and challenges. *AIP Advances*, 15(2):025112, 2025. doi: [10.1063/5.0255451](https://doi.org/10.1063/5.0255451).
- [26] H.O. Ahmed and A.K. Nandi. Convolutional-transformer model with long-range temporal dependencies for bearing fault diagnosis using vibration signals. *Machines*, 11(7):746, 2023. doi: [10.3390/machines11070746](https://doi.org/10.3390/machines11070746).

- [27] P. Luo and Z. Liu. Unsupervised bearing fault diagnosis using masked self-supervised learning and Swin transformer. *Machines*, 13(9):792, 2025. doi: [10.3390/machines13090792](https://doi.org/10.3390/machines13090792).
- [28] Q. Chen, F. Zhang, Y. Wang, Q. Yu, G. Lang, and L. Zeng. Bearing fault diagnosis based on efficient cross-space multiscale CNN transformer parallelism. *Scientific Reports*, 15(1):12344, 2025. doi: [10.1038/s41598-025-95895-x](https://doi.org/10.1038/s41598-025-95895-x).
- [29] M.F. Siddique, F. Saleem, M. Umar, C.H. Kim, and J.M. Kim. A hybrid deep learning approach for bearing fault diagnosis using continuous wavelet transform and attention-enhanced spatiotemporal feature extraction. *Sensors*, 25(9):2712, 2025. doi: [10.3390/s25092712](https://doi.org/10.3390/s25092712).
- [30] M. Sehri and P. Dumond. University of Ottawa rolling-element dataset – vibration and acoustic faults under constant load and speed conditions (UORED-VAFCLS). *Mendeley Data*, V2, 2023. doi: [10.17632/y2px5tg92h.2](https://doi.org/10.17632/y2px5tg92h.2).