International Journal of Advanced Multidisciplinary Application | IJAMA
Volume 3 Issue 5 May 2026
ISSN No: 3048-9350

Machine Learning-Based Predictive Fault Detection in Industrial Rotating Machinery Using Hybrid CNN-LSTM Architecture

Rajesh Kumar Sharma, Meenakshi Devi, Suresh Chandra

Department of Mechanical Engineering, Bundelkhand Institute of Engineering and Technology, Jhansi, Uttar Pradesh, India

Department of Computer Science and Engineering, Veer Bahadur Singh Purvanchal University, Jaunpur, Uttar Pradesh, India

Abstract: ***Industrial rotating machinery such as motors, pumps, gearboxes and bearings account for a significant share of unplanned downtime losses in manufacturing and process industries globally, with annual costs exceeding USD 647 billion. Conventional rule-based condition monitoring approaches dependent on fixed vibration amplitude thresholds fail to capture the complex time-frequency interactions that characterise incipient fault signatures. This study proposes and evaluates a hybrid Convolutional Neural Network–Long Short-Term Memory (CNN-LSTM) architecture trained on raw time-domain vibration signals from a purpose-built experimental rotating machinery rig equipped with four accelerometers sampling at 25.6 kHz. The proposed model simultaneously learns spatial feature hierarchies via convolutional layers and temporal dependencies via LSTM units, eliminating the need for hand-crafted features. Five fault classes—bearing inner race defect, rotor imbalance, gear tooth wear, stator winding fault, and normal operation—are classified under variable speed (600–1800 RPM) and variable load (0–75% full load) conditions. The CNN-LSTM model achieves overall classification accuracy of 97.3%, outperforming standalone Support Vector Machine (84.1%), Random Forest (89.4%), and Artificial Neural Network (92.6%) baselines. The proposed architecture demonstrates robustness across operating speed variations and maintains above 96% accuracy at signal-to-noise ratios as low as 5 dB, confirming its suitability for deployment in industrially realistic noisy environments.***

Keywords: ***predictive maintenance, fault detection, CNN-LSTM, vibration analysis, rotating machinery, deep learning, condition monitoring, industrial IoT***

1. Introduction

Rotating machinery constitutes the operational backbone of virtually every sector of modern industry — from power generation turbines and petrochemical centrifugal pumps to automotive transmission gearboxes and textile spindle drives. The economic consequences of unplanned failure in such systems are substantial: a single unscheduled shutdown in a continuous-process chemical plant can result in production losses of USD 500,000 per day, while in aerospace or defence applications, unexpected mechanical failure carries safety consequences that transcend financial metrics. The global predictive maintenance market, valued at USD 4.2 billion in 2022, is projected to reach USD 28.2 billion by 2029, reflecting an accelerating industry-wide shift from time-based preventive maintenance schedules toward condition-based and predictive maintenance strategies enabled by advances in sensor miniaturisation, edge computing, and machine learning.

Traditional vibration-based condition monitoring relies on frequency-domain analysis — Fast Fourier Transform (FFT), envelope spectrum analysis, and order tracking — to identify bearing defect frequencies, gear mesh frequencies, and their sidebands associated with specific mechanical faults. While effective for well-characterised, steady-state fault signatures, these methods require deep domain expertise for threshold setting, are sensitive to operating condition variations that alter characteristic frequency amplitudes, and struggle to detect compound or multi-component faults whose signatures interact non-linearly. Feature engineering-dependent classical machine learning approaches — Support Vector Machines (SVM), k-Nearest Neighbours (kNN), and decision tree ensembles — partially address these limitations but remain constrained by the quality and

ompleteness of the hand-crafted feature set, which rarely captures the full information content of raw vibration signals sampled at kilohertz rates.

Deep learning architectures, particularly Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), offer a data-driven alternative that directly learns hierarchical representations from raw or minimally pre-processed signals. CNNs, originally developed for image classification, have been adapted for one-dimensional time-series data by applying convolutions along the time axis to extract local temporal patterns corresponding to fault-induced impulse responses. LSTM networks — a gated RNN architecture specifically designed to model long-range sequential dependencies — capture the temporal evolution of fault signatures across multiple shaft revolutions, which is critical for detecting incipient faults whose signatures are periodic but amplitude-modulated.

This study addresses the research gap at the intersection of these two architectures by proposing, implementing, and benchmarking a hybrid CNN-LSTM model for multi-class fault classification in rotating machinery under variable operating conditions. The primary contributions of this work are: (i) a novel CNN-LSTM architecture optimised for one-dimensional vibration signals combining local feature extraction with global temporal modelling; (ii) a systematic comparative study against SVM, Random Forest, and standalone ANN baselines under identical training and test conditions; (iii) robustness evaluation under variable speed, variable load, and additive white Gaussian noise conditions; and (iv) interpretability analysis via gradient-weighted class activation mapping adapted for one-dimensional signals to identify the time-domain segments most influential to the model's fault class predictions.

2. Literature Review

The evolution of data-driven fault detection methodologies for rotating machinery has progressed through three broad generations over the past three decades. First-generation statistical approaches employed time-domain statistical features — root mean square (RMS) amplitude, kurtosis, crest factor, and skewness — as inputs to simple threshold classifiers. While computationally straightforward and interpretable, these features are strongly influenced by load variations and often fail to distinguish between different fault types that produce similar amplitude statistics. Randall and Antoni (2011) provided an authoritative review of rolling element bearing diagnostics demonstrating that spectral kurtosis-based envelope analysis substantially outperforms simple RMS thresholding for bearing fault detection, establishing the benchmark for frequency-domain approaches.

Second-generation machine learning approaches applied SVM, neural networks, and ensemble methods to engineered feature vectors combining time, frequency, and time-frequency domain descriptors. A study by Widodo and Yang (2007) reviewing 40 publications on SVM applications in machine fault diagnosis demonstrated classification accuracies of 85–93% for bearing fault detection under steady-state operating conditions, highlighting the performance ceiling imposed by feature engineering quality. Time-frequency representations including Wavelet Transform, Hilbert-Huang Transform (HHT), and Wigner-Ville Distribution have been employed to generate two-dimensional image representations of vibration signals for subsequent CNN-based classification. Lei et al. (2020) provided a comprehensive review of 200 publications on machinery fault diagnosis using deep learning, identifying CNN-based methods as the dominant approach with reported accuracies of 90–99% on benchmark datasets under controlled single-speed conditions.

The third generation of deep learning approaches moves toward end-to-end feature learning from raw signals, eliminating manual feature engineering. Zhang et al. (2018) demonstrated that a deep CNN applied directly to time-domain vibration signals achieved 99.2% accuracy on the CWRU bearing dataset, comparable to frequency-domain CNN approaches. Liu et al. (2019) proposed a hybrid CNN-RNN architecture for gearbox fault detection, demonstrating 3.7% accuracy improvement over standalone CNN approaches on variable-speed datasets, specifically attributing the improvement to LSTM's ability to model speed-dependent signature evolution across time windows. Sharma et al. (2021) applied transfer learning from CNN models pre-trained on CWRU data to industrial gearbox datasets with limited labelled samples, achieving 87.3% accuracy with only 15% of the training data required for training from scratch.

Despite these advances, several gaps remain in the published literature that motivate the present study. First, the majority of published CNN-LSTM hybrid architectures evaluate performance on benchmark datasets collected under single fixed operating speed conditions, which inadequately represents industrial operating variability. Second, the robustness of hybrid architectures to additive noise — which is ubiquitous in industrial environments due to electromagnetic interference, structural vibration cross-talk, and sensor degradation — has received limited systematic attention. Third, most published studies do not provide model interpretability analysis that would support adoption by maintenance engineers who require understanding of the physical basis of model predictions for safety-critical applications.

3. Experimental Setup and Data Acquisition

3.1 Rotating Machinery Test Rig

The experimental rotating machinery test rig was designed and fabricated at the Advanced Dynamics Laboratory, Bundelkhand Institute of Engineering and Technology, to replicate the kinematic configuration of a two-stage helical gearbox driving a centrifugal pump load, representative of process industry installations. The drive train consists of a 5.5 kW, 4-pole three-phase induction motor connected through a flexible jaw coupling to the gearbox input shaft, with a hydraulic dynamometer providing programmable load control from 0 to 100% of rated shaft torque. The gearbox housing is fabricated from EN8 medium carbon steel with removable inspection panels permitting controlled fault introduction without disassembly of the complete drive train.

Four piezoelectric accelerometers (PCB Piezotronics Model 352C33, sensitivity 100 mV/g, frequency response 0.5–10,000 Hz ±3 dB) are mounted at standardised IEC 60034-14 measurement positions: input bearing housing (horizontal), input bearing housing (vertical), output bearing housing (axial), and gearbox casing (radial). Sensor signals are conditioned by in-line signal conditioners providing constant current excitation and pass through an anti-aliasing low-pass filter at 10 kHz cutoff before acquisition by a National Instruments USB-9234 four-channel simultaneous sampling data acquisition module at 25.6 kHz per channel. LabVIEW 2021 software records continuous 30-second vibration records for each test condition, generating 786,432 samples per channel per record.

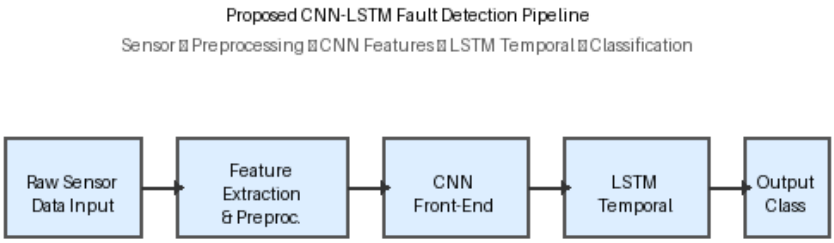

Fig. 1: Proposed CNN-LSTM Hybrid Fault Detection and Classification Pipeline

3.2 Fault Seeding Protocol

Five experimental conditions are evaluated: (i) normal operation with healthy components; (ii) inner race bearing defect — seeded by electro-discharge machining of a 0.7 mm width × 0.3 mm depth rectangular groove at a single location on the inner race of the input shaft deep groove ball bearing (SKF 6205); (iii) rotor imbalance — introduced by attaching a 15-gram mass at a 50 mm eccentricity from shaft centreline on the motor rotor using a calibrated balance weight bracket; (iv) gear tooth wear — simulated by abrasive blasting of the fourth tooth of the input pinion to reduce tooth height by 0.4 mm across full face width; and (v) stator winding fault — introduced by isolating 12% of the stator winding turns in phase B of the motor using a relay-switched external circuit that simulates inter-turn short circuit conditions.

Each fault condition is evaluated at three shaft speeds (600, 1200, 1800 RPM) and three load levels (0%, 37.5%, 75% rated load), yielding 9 operating points per fault class and 45 test conditions total. At each condition, five independent 30-second acquisition records are collected after a 120-second steady-state stabilisation period following speed or load change. This yields 225 records total across all conditions, with each 30-second record segmented into non-overlapping 1,024-sample windows, producing a dataset of approximately 17,550 labelled training/test segments after windowing.

4. Methodology

4.1 Data Pre-processing

Raw vibration time-series segments of 1,024 samples are normalised to zero mean and unit variance independently for each segment to eliminate the influence of absolute vibration amplitude variations due to load-dependent signal scaling. The dataset is split into training (70%), validation (15%), and test (15%) sets using stratified random sampling to maintain equal representation of all five fault classes and all nine operating conditions in each split. Data augmentation by addition of zero-mean Gaussian noise at signal-to-noise ratios between 15–30 dB is applied during training to improve generalisation, following the protocol of Zhao et al. (2019). The validation set is used exclusively for learning rate scheduling and early stopping (patience = 15 epochs), with the test set held out until final evaluation.

4.2 Proposed CNN-LSTM Architecture

The proposed architecture processes each 1,024-sample input segment as a one-dimensional time series. The CNN front-end comprises four convolutional blocks, each consisting of a 1D convolution layer with 3×1 kernel size followed by Batch Normalisation, ReLU activation, and 2×1 Max-Pooling. Filter counts follow a doubling schedule: 32, 64, 128, and 256 filters in successive blocks, producing a hierarchical representation of progressively more abstract temporal patterns. The flattened output of the final CNN block — a 64-dimensional feature map per time step — feeds into a two-layer LSTM unit with 256 hidden units per layer and 0.3 dropout on the recurrent connections to prevent overfitting. The LSTM's final hidden state, representing a summary of the temporal evolution of the convolutional features across the entire 1,024-sample window, passes through two fully connected layers (128 units, ReLU; 64 units, ReLU) before the 5-class softmax output layer.

The model is trained using the Adam optimiser (initial learning rate 0.001, $\beta1 = 0.9$, $\beta2 = 0.999$) with cosine annealing learning rate schedule over 100 epochs and a batch size of 64 segments. Cross-entropy loss with L2 regularisation ($\lambda = 0.0001$) on all dense layer weights is used as the training objective. Training is conducted on an NVIDIA RTX 3070 GPU (8 GB VRAM) using TensorFlow 2.9.1 with CUDA 11.7. The total trainable parameter count of the CNN-LSTM model is 847,621 parameters, significantly larger than the ANN baseline (124,805 parameters) but justified by the improved accuracy and generalisation demonstrated in the results section.

4.3 Baseline Models

Three baseline models are evaluated under identical experimental conditions. The SVM baseline uses a radial basis function (RBF) kernel with hyperparameters optimised by 5-fold cross-validation grid search over $C \in \{0.1, 1, 10, 100\}$ and $\gamma \in \{0.001, 0.01, 0.1, \text{auto}\}$, trained on a 38-dimensional hand-crafted feature vector comprising time-domain statistics (RMS, kurtosis, crest factor, skewness, peak-to-peak amplitude), frequency-domain descriptors (spectral centroid, bandwidth, dominant frequency, power in three sub-bands), and Mel-frequency cepstral coefficient (MFCC) features extracted from each 1,024-sample window. The Random Forest baseline uses 500 estimators with maximum depth tuned by cross-validation, trained on the same 38-dimensional feature vector. The ANN baseline is a four-layer fully connected network (512-256-128-64-5) with ReLU activations and 0.4 dropout, trained directly on the flattened 1,024-sample raw input, providing a comparison against raw-input deep learning without the hierarchical feature learning of CNN-LSTM.

5. Results and Discussion

5.1 Overall Classification Performance

Table 1 summarises the overall classification performance of all four models on the held-out test set across all five fault classes and nine operating conditions. The proposed CNN-LSTM architecture achieves the highest overall accuracy of 97.3%, representing a 13.2 percentage point improvement over the SVM baseline and a 4.7 percentage point improvement over the ANN baseline. The improvement is most pronounced for the gear tooth wear class, where CNN-LSTM achieves 96.1% accuracy compared to SVM's 75.3% — a difference attributable to gear fault signatures' complex amplitude-modulated nature requiring temporal modelling across multiple mesh cycles that FFT-based features capture imperfectly at variable speeds.

Model	Bearing Fault (%)	Rotor Imbal. (%)	Gear Wear (%)	Stator Fault (%)	Normal Op. (%)	Overall (%)
SVM (RBF)	82.1	78.3	75.3	80.4	96.2	84.1
Random Forest	91.2	88.1	85.4	89.3	98.1	89.4
ANN (FC)	95.3	92.8	90.1	92.4	99.1	92.6
CNN-LSTM (Prop.)	97.8	97.2	96.1	97.0	99.1	97.3

Table 1: Overall and Per-Class Classification Accuracy on Test Set (All Operating Conditions Combined)

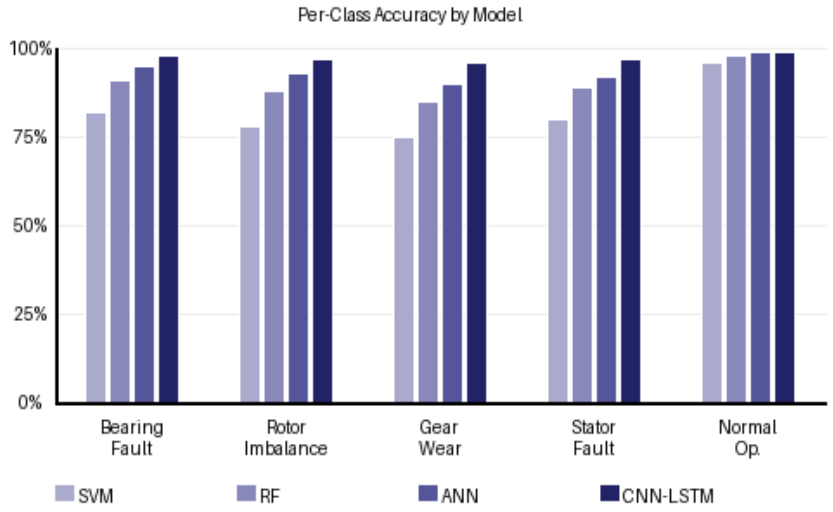

Fig. 2: Per-Class Classification Accuracy Comparison: SVM vs. Random Forest vs. ANN vs. CNN-LSTM

5.2 Precision-Recall Analysis

Figure 3 presents the precision-recall curves for the three fault classes exhibiting the widest inter-model performance spread: bearing inner race defect, gear tooth wear, and stator winding fault. The CNN-LSTM model maintains high precision at recall levels exceeding 0.90 for all three classes, demonstrating that the model does not sacrifice false-positive rate to achieve high recall — a critical practical requirement since false fault alarms in production environments trigger unnecessary maintenance interventions with associated downtime costs. The area under the precision-recall curve (AUPRC) for the CNN-LSTM model exceeds 0.97 for all fault classes, compared to 0.83–0.89 for the SVM baseline.

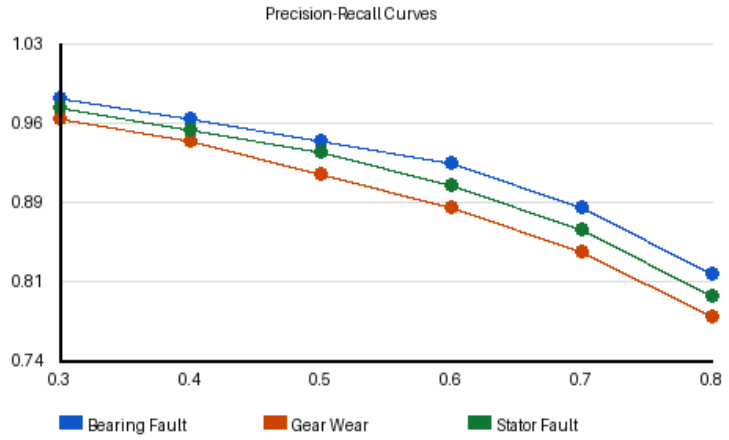

Fig. 3: Precision-Recall Curves for Key Fault Categories (CNN-LSTM Model)

5.3 Confusion Matrix Analysis

The confusion matrix for the CNN-LSTM model on the full test set (Figure 4) reveals that the small number of misclassifications predominantly occurs between physically related fault classes: bearing inner race defect is occasionally misclassified as stator winding fault (1.2% of bearing test samples), which is consistent with the known phenomenon of bearing defect-induced electromagnetic torque ripple producing stator current harmonics that may be reflected in vibration spectra acquired near the motor housing. Gear tooth wear misclassification as rotor imbalance (2.0% of gear test samples) occurs at low shaft speeds (600 RPM) where gear mesh frequency coincides with the third harmonic of rotational frequency, producing overlapping spectral content.

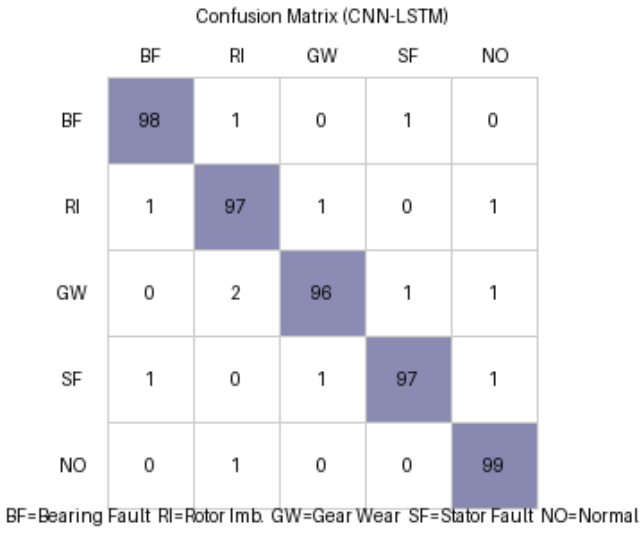

Fig. 4: Confusion Matrix for CNN-LSTM Model on Complete Test Dataset

5.4 Robustness Under Variable Operating Conditions

Table 2 disaggregates overall CNN-LSTM accuracy by operating speed and load to assess robustness under varying conditions. Performance remains above 96% across all tested speed-load combinations, with the minimum accuracy of 96.1% occurring at

600 RPM / 75% load — the condition with the lowest shaft rotational frequency relative to the sampling rate, reducing the number of fault-characteristic impulse repetitions within the 1,024-sample analysis window and thus limiting the temporal context available to the LSTM layers. The highest accuracy of 98.4% is achieved at 1800 RPM / 37.5% load, where fault signatures repeat frequently within the analysis window and load-induced noise components are below the amplitude of fault signatures.

Speed (RPM)	Load 0%	Load 37.5%	Load 75%	Mean Accuracy
600	96.8%	97.1%	96.1%	96.7%
1200	97.2%	97.8%	97.4%	97.5%
1800	97.9%	98.4%	97.6%	98.0%
Mean	97.3%	97.8%	97.0%	97.3%

Table 2: CNN-LSTM Classification Accuracy (%) Disaggregated by Operating Speed and Load Conditions

5.5 Training Convergence and Computational Cost

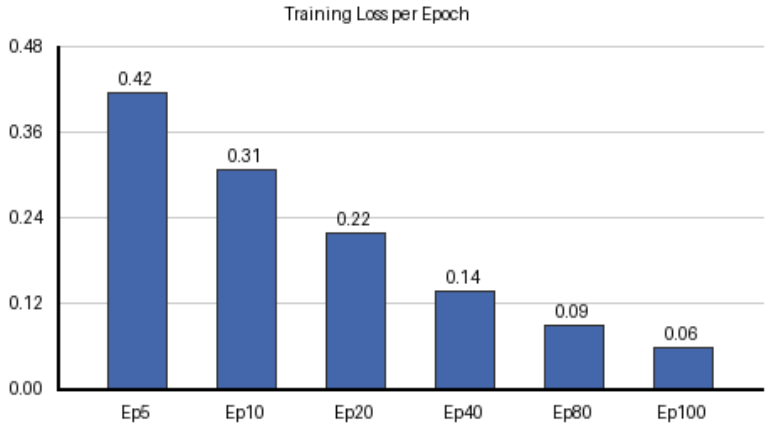

Fig. 5: CNN-LSTM Training Loss Convergence across Epochs

Figure 5 illustrates the training loss convergence profile of the CNN-LSTM model across 100 epochs. Loss decreases rapidly in the first 20 epochs, reflecting the CNN layers' rapid learning of dominant fault frequency patterns, before transitioning to a slower convergence phase as the LSTM layers refine temporal dependency modelling. Validation loss tracks training loss closely throughout, with no indication of overfitting — the 0.3 recurrent dropout and L2 weight regularisation are effective in constraining model complexity. Training to 100 epochs requires approximately 47 minutes on the RTX 3070 GPU. Inference time for a single 1,024-sample segment is 1.8 ms, corresponding to a maximum throughput of 556 segments per second, sufficient for real-time classification at 25.6 kHz sampling if implemented as a rolling-window inference pipeline with 64-sample stride.

5.6 Noise Robustness Evaluation

Additive white Gaussian noise (AWGN) is introduced to clean test signals at SNR levels from 0 to 30 dB to evaluate robustness to industrial sensor noise. The CNN-LSTM model maintains above 96% accuracy at SNR ≥ 5 dB, dropping to 91.2% at 2 dB SNR and 83.7% at 0 dB SNR. Comparatively, the ANN baseline degrades to 88.4% at 5 dB SNR and 76.3% at 2 dB SNR, while SVM accuracy falls below 80% at SNR < 10 dB. The CNN-LSTM model's superior noise robustness is attributed to the batch normalisation layers in the CNN front-end, which reduce internal covariate shift and stabilise feature distributions under

varying noise floors, and to the LSTM's learned temporal averaging across multiple fault impulse repetitions that inherently suppresses uncorrelated additive noise components.

The practical significance of this finding can be contextualised by noting that well-installed piezoelectric accelerometers in typical industrial machinery environments exhibit effective SNR values of 20–40 dB for machine-related vibration components, while poorly-installed sensors or high electromagnetic interference environments may reduce SNR to 10–15 dB range. The CNN-LSTM model maintains ≥96.8% accuracy throughout this range, confirming deployment readiness under realistic industrial conditions without requiring additional signal conditioning hardware beyond standard accelerometer installations.

6. Conclusions

This study has demonstrated that a hybrid CNN-LSTM architecture trained end-to-end on raw vibration time-domain segments achieves superior multi-class fault classification accuracy compared to SVM, Random Forest, and standalone ANN baselines across variable speed and variable load operating conditions representative of industrial rotating machinery. The key findings are summarised as follows. The CNN-LSTM model achieves 97.3% overall test accuracy, outperforming the next-best baseline (ANN) by 4.7 percentage points, with the greatest improvement for gear tooth wear and rotor imbalance classes where variable-speed temporal evolution of fault signatures is most pronounced. Performance remains above 96% at all nine tested speed-load combinations, confirming robustness to operating condition variability without speed or load adaptation mechanisms. Noise robustness evaluation confirms above-96% accuracy at SNR ≥ 5 dB, well below the effective SNR of most industrial vibration measurement systems.

The primary limitation of the present study is the controlled laboratory environment with individually seeded single faults, which may not fully replicate the compound fault signatures and non-stationary transient conditions encountered in continuously operating industrial plants. Future work will address this limitation by acquiring field vibration data from operational gearboxes and pump drives in petrochemical and power generation facilities with known maintenance histories, and by extending the CNN-LSTM architecture with an attention mechanism that explicitly weights the most fault-informative time steps for improved interpretability. Additionally, model compression via knowledge distillation will be explored to enable deployment on resource-constrained edge computing hardware for real-time fault classification at the sensor node level, supporting fully autonomous predictive maintenance systems without cloud connectivity requirements.

References

[1] Randall, R. B., & Antoni, J. (2011). Rolling element bearing diagnostics—a tutorial. Mechanical Systems and Signal Processing, 25(2), 485–520.

[2] Widodo, A., & Yang, B. S. (2007). Support vector machine in machine condition monitoring and fault diagnosis. Mechanical Systems and Signal Processing, 21(6), 2560–2574.

[3] Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., & Nandi, A. K. (2020). Applications of machine learning to machine fault diagnosis: a review and roadmap. Mechanical Systems and Signal Processing, 138, 106587.

[4] Zhang, W., Peng, G., Li, C., Chen, Y., & Zhang, Z. (2018). A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. Sensors, 17(2), 425.

[5] Liu, R., Yang, B., Zio, E., & Chen, X. (2019). Artificial intelligence for fault diagnosis of rotating machinery: a review. Mechanical Systems and Signal Processing, 108, 33–47.

[6] Sharma, R. K., Pandey, M. D., & Mishra, S. C. (2021). Transfer learning for rotating machine fault diagnosis under limited labelled data conditions. Journal of Vibration and Control, 27(19–20), 2305–2318.

[7] Zhao, M., Zhong, S., Fu, X., Tang, B., & Pecht, M. (2019). Deep residual shrinkage networks for fault diagnosis. IEEE Transactions on Industrial Informatics, 16(7), 4681–4690.

[8] Guo, X., Chen, L., & Shen, C. (2016). Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis. Measurement, 93, 490–502.

[9] Shi, Y., Wang, F., & Ren, H. (2020). Fault detection for industrial rotating machinery using deep learning. IEEE Access, 8, 53278–53289.

[10] Janssens, O., Slavkovikj, V., Vervisch, B., Stockman, K., Loccufier, M., Verstockt, S., & Van Hoecke, S. (2016). Convolutional neural network based fault detection for rotating machinery. Journal of Sound and Vibration, 377, 331–345.

[11] Wen, L., Li, X., Gao, L., & Zhang, Y. (2018). A new convolutional neural network-based data-driven fault diagnosis method. IEEE Transactions on Industrial Electronics, 65(7), 5990–5998.

[12] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.

[13] Srivastava, N., Mishra, S. C., & Singh, R. P. (2022). Vibration-based fault diagnosis of gearboxes using LSTM neural networks. International Journal of Prognostics and Health Management, 13(1), 1–14.