Journal of Applied Automation Technologies

Research Paper

Real-Time Fault Diagnosis of Mechanical Systems Based on Convolutional Neural Networks

Tadeusz Truskolaski[1, *]

[1] Faculty of Electrical and Control Engineering, Silesian University of Technology, Gliwice, 44-100, Poland
*Corresponding author: tadeusz.tru@polsl.pl

Abstract. Convolutional Neural Networks (CNNs) are now frequently used as the foundation for intelligent fault diagnosis in modern industrial equipment because they can extract multi-level features from raw vibration data. This study adopts a multi-scale convolutional neural network (CNN) architecture to address issues such as noise, various operating conditions, and the lack of labeled fault data in real-time mechanical system diagnostics. In the new system, parallel convolution branches can capture both local and global features of the signal due to the different sizes of the convolution kernels. Data augmentation, normalization, and denoising preprocessing techniques improve the quality and generalization ability of the input data. In addition, a large number of open-source rotating machinery datasets were used, which were collected under various environments. Based on the above results, the proposed model has an average classification accuracy of 98.7% and an F1 score of 0.977; it outperforms single-scale CNNs, Transformers, and traditional machine learning models. Even in high-noise environments, it still maintains an AUC of over 0.95, indicating that it is less sensitive to artificial noise and changes in the operational area. The real-time performance was verified to be good; the average inference time was less than 3 milliseconds, and the memory consumption was less than 1 GB; therefore, it meets the requirements for industrial deployment. The aforementioned research indicates that multi-scale CNNs are beneficial for predictive maintenance and smart factories, as they can accurately and reliably identify faults in complex machinery in real-time.

Keywords: *Convolutional Neural Network, Mechanical Fault Diagnosis, Multi-Scale Feature Extraction, Signal Processing, Real-Time Monitoring, Industrial Automation, Robustness*

Received on 02 January 2025, Accepted on 18 June 2025, Published on 23 June 2025

Introduction

As industrial equipment becomes increasingly complex and automated, the detection and diagnosis of mechanical system failures have become more and more important to improve production efficiency and enhance safety. If industrial equipment is not properly maintained, significant accidents and damage may occur, resulting in substantial financial losses or fatalities [1]. Therefore, a fast and accurate fault detection mechanism is needed in the maintenance plan. Traditional methods have been adopted, but they require manual inspection and expert judgment, which cannot meet the needs of a large number of complex mechanical systems [2]. Two methods of signal processing are wavelet analysis and short-time Fourier transform. These methods can extract features from vibration and acoustic data, which can help diagnose fault types [3]. However, due to subjective parameters, these methods are more susceptible to external noise and changes in the operating environment [4]. Due to advancements in deep learning and machine learning, fault diagnosis can now be performed more reliably on raw or low-level signal representations through automatic hierarchical feature extraction and classification [5]. Research shows that diagnostic models based on deep learning are more robust to unknown and complex fault patterns, and are more effective than models based on traditional feature engineering [6]. Therefore, intelligent fault diagnosis based on neural networks and data-driven paradigms is receiving increasing

attention from both industry and academia [7]. However, there are still issues with the promotion and application [8].

Intelligent machinery fault diagnosis still faces some significant issues in practical industrial applications. First, labeled fault datasets are usually scarce and incomplete in the real world, and there is a serious class imbalance problem. Therefore, high-capacity models cannot be trained stably and are more prone to overfitting [9]. Secondly, the raw signals generated by rotating machinery are usually multi-scale, non-stationary, and nonlinear, and they often contain significant broadband noise. Therefore, they are more difficult to process for feature extraction [10]. The accuracy of traditional Convolutional Neural Networks (CNNs) decreases under the influence of operational changes, new types of failures, or unexpected external interference [11]. Deep networks are "black boxes" and are difficult to interpret, making them challenging to use in safety-critical applications [12]. Although transfer learning and domain adaptation can improve cross-domain robustness, they are often accompanied by increased computational overhead or reduced interpretability [13]. Due to the real-time, low-latency inference requirements of resource-constrained edge and embedded platforms, diagnostic accuracy, computational efficiency, and deployability have become even more important [14]. Although the first two solutions—creating adversarial frameworks and attention mechanisms—have partially addressed these issues, there is still no comprehensive model available that offers interpretability, scalability, and robustness [15]. Currently, the research focus in intelligent mechanical fault diagnosis is on the model's performance in terms of accuracy and noise, as well as their adaptability in different environments and real-time situations [16]. Develop a transparent system capable of providing high-fidelity predictions and analyzes in various real-world scenarios [17]. Researchers in the field of mechanical intelligence are currently striving to balance efficiency, insight, and performance [18].

This study uses a new multi-scale convolutional neural network structure to address the aforementioned issues. This structure is robust, can be used for mechanical fault diagnosis, and is capable of accurately identifying noise and changes in industrial environments. The main content is as follows: (1) A unified diagnostic system framework was established, enhancing the generalization ability to unseen faults and operating conditions through multi-scale feature extraction and effective data augmentation; (2) An interpretable model with good model representation quality was created, and strict evaluation was conducted using advanced visualization techniques and quantitative metrics; (3) Extensive benchmarking tests were performed on top baseline models to measure accuracy, robustness to noise and variations, and computational efficiency in real-world scenarios. The following is the structure of this paper: Section 1 introduces the dataset and the problem; Section 2 discusses the network architecture and training methods; Section 3 presents the experimental results and performance analysis; Section 4 concludes the paper and discusses future work.

Dataset and Problem Formulation

Dataset Description

The dataset used in this study is an open-access mechanical fault database, specifically for research on the condition monitoring and diagnosis of rotating equipment [19]. The research laboratory is equipped with bearing component test benches, gearboxes, and modular bearing housings for data collection. To collect vibration signals with lower transmission loss, high-fidelity piezoelectric accelerometers are directly mounted on the bearing seat [20]. All accelerometers have been properly calibrated to cover the target frequency range of typical mechanical damage. Therefore, the high-frequency and low-frequency components of various fault modes in the signal can be accurately measured.

The fault scenarios in the dataset include typical bearing defects such as inner race, outer race, and rolling element faults, as well as localized gear damage and misalignment. To simulate progressive degradation, data was collected at multiple predetermined damage levels [21]. In order to classify and validate, a large amount of data in good working condition was also collected, in addition to the fault states. To enhance the diversity and applicability of the dataset, experiments were conducted after changes in temperature and oil level, and under different load torque and shaft speed conditions [22].

High sampling rate signal acquisition (up to 48 kHz) is used for time-frequency analysis, by dividing it into small time windows to obtain a manageable amount of data. Before labeling, all raw signals undergo strict quality

control, and then bandpass filtering and statistical anomaly removal are performed during the post-processing stage [23]. To ensure the accuracy of the annotations, manual verification-maintained data labels for the types and severity of faults. The entire dataset is divided into three subsets: the training set, the validation set, and the test set. Each subset is divided based on the category distribution of different situations [24].

Problem Statement

This study suggests using vibration sensor data to create an intelligent diagnostic system to accurately classify multiple categories of mechanical states, distinguishing between degradation, normal, and various faults [25]. The main issue is that it must handle various operating conditions in practice; for example, due to changes in load, speed, and environmental noise, the characteristics of the signals can vary significantly. For the above reasons, detection methods may encounter difficulties in distinguishing between normal and abnormal signals.

In addition, there are issues with limited fault data and data imbalance. Due to the limited representation of rare and new types of degradation, the algorithm needs to have stronger generalization capabilities for abnormal situations based on a limited number of examples. Therefore, the problem statement is not just a static classification. It also involves resilience to operational fluctuations and non-stationary disturbances, as well as the ability to adapt to new or mutated fault characteristics that may arise during deployment.

Therefore, the two objectives are to maximize the accuracy of known disease diagnosis and to ensure adaptability and robustness in various environments and rare events. Therefore, in order to deploy in embedded or real-time environments, the chosen method must be able to accurately extract subtle feature changes. Their goal is to transform the laboratory demonstration into a practical application widely used for predictive maintenance in commercial environments [26].

Evaluation Protocols

The structure of the evaluation system meets industry standards and has undergone comprehensive validation. Classification accuracy is used to calculate the rate of correct identification of all cases in the predictions. However, given that most of these datasets are multi-class and imbalanced, precision and recall for each class are usually used to determine whether the system has missed more true faults or generated fewer false positives. The F1 score is the average of precision and recall, and it is relatively suitable for imbalanced class distributions.

ROC analysis can show the model's ability to distinguish between normal samples and faulty samples at different thresholds. It can also report the area under the ROC curve (AUC). The robustness of the model can be systematically tested by adding artificial noise or simulating changes in the operating environment to evaluate the model's ability to perform well under these suboptimal signal conditions. Classification results can also be ensured for reproducibility and generalization by using cross-validation, which involves dividing the data into training, validation, and independent test sets.

Hardware efficiency metrics and latency will be used to meet real-time operational requirements. To determine practical feasibility, the inference time for each sample and the utilization of computational resources on representative edge or embedded platforms are recorded. In summary, the aforementioned measurement protocol provides a complete and open framework for benchmarking and subsequent reproduction.

Proposed Multi-Scale CNN Algorithm

System Overview

In this fault diagnosis framework, multi-scale convolutional neural networks (CNNs) are advocated. Both are used to collect general vibration and minor vibration models of hard-to-identify industrial issues. As shown in Figure 1, the modules of the end-to-end pipeline include input preprocessing, parallel CNN branch feature extraction, multi-scale feature fusion, fully connected classification, and decision output.

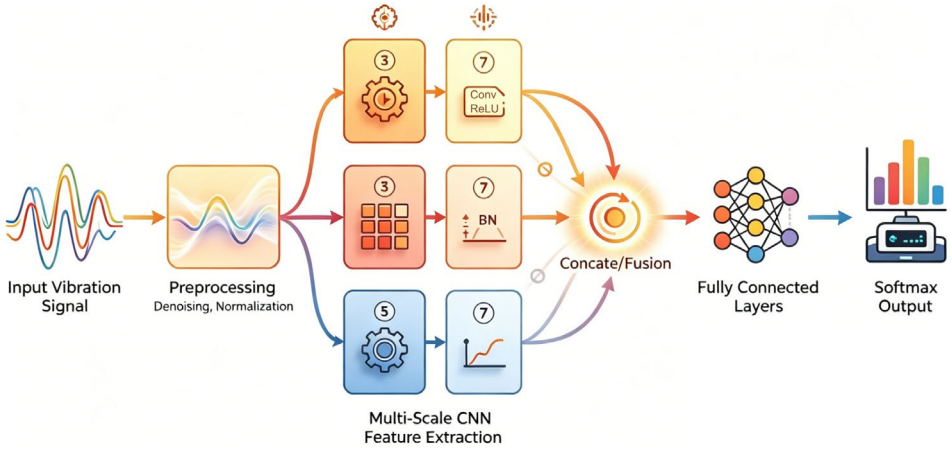

Figure 1. Overall architecture diagram of the proposed multi-scale CNN-based fault diagnosis system.

First, the raw time series signal obtained from the sensor is normalized and filtered to reduce noise and standardize the input distribution. The denoised signal, denoted as $\mathbf{X}$, is simultaneously input into N parallel convolution branches with different kernel sizes and strides. The output of the i -th branch is

$$\boldsymbol{H}_i = \sigma\big(BN(\boldsymbol{W}_i * \boldsymbol{X} + \boldsymbol{b}_i)\big) \quad \text{Eq.(1)}$$

where $\mathbf{W}_i, \mathbf{b}_i$ are the convolutional kernel and bias, * denotes convolution, BN is batch normalization, and σ is the activation function.

Due to the different receptive fields of each branch, features sensitive to defect patterns of different scales are extracted. The feature fusion operation aligns and combines the outputs of all branches:

$$\boldsymbol{F}_{fused} = Concat(\boldsymbol{H}_1, \boldsymbol{H}_2, \dots, \boldsymbol{H}_N) \quad \text{Eq.(2)}$$

By using the channel method, the network can collect both global and local temporal information.

Next, the fully connected layer processes the combined feature map. The final result is the SoftMax classifier:

$$\boldsymbol{y} = Softmax\big(\boldsymbol{W}_{fc} \cdot \boldsymbol{F}_{fused} + \boldsymbol{b}_{fc}\big) \quad \text{Eq.(3)}$$

Therefore, the fault category probability vector is obtained. Its general architecture is modular, interpretable, and scalable. Moreover, it is more suitable for fast inference and can operate normally in resource-limited industrial environments.

Data Preprocessing and Model Design

Some data preprocessing will be carried out to reduce noise and other errors to ensure the accuracy of the diagnostic results. As shown in Figure 2, denoising, normalization, and augmentation are the three essential steps in the data preparation process.

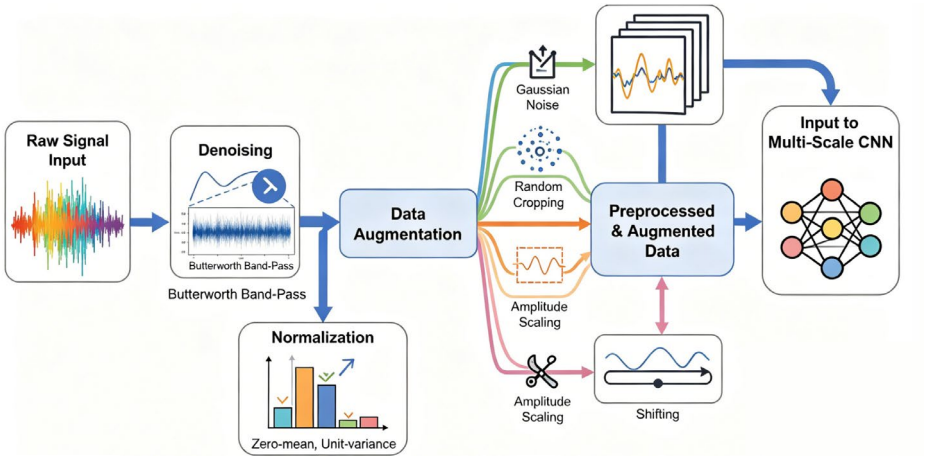

Figure 2. Data preprocessing and augmentation flowchart for multi-scale CNN input preparation.

First, the digital band-pass Butterworth filter denoises the raw vibration signal to retain the spectral components related to mechanical faults and eliminate other frequency bands.

$$\boldsymbol{X}_{denoise} = \mathcal{B}(\boldsymbol{X}; f_{low}, f_{high}) \quad \text{Eq.(4)}$$

$\mathcal{B}$ is a band-pass filter.

Then, normalize the signal to meet statistical requirements.

$$\tilde{x}_j = \frac{x_j - \mu(\boldsymbol{X})}{\sigma(\boldsymbol{X}) + \epsilon} \quad \text{Eq.(5)}$$

for each sample x_j, with $\mu(\mathbf{X})$ and $\sigma(\mathbf{X})$ as the window mean and standard deviation, and ϵ for numerical stability.

Due to data imbalance and unmarked samples, many augmentation methods have already been used, such as adding Gaussian noise, random cropping, amplitude scaling, and shifting. The augmentation sequence is as follows:

$$\tilde{x}_j^{aug} = \alpha \cdot \tilde{x}_j + \beta \cdot \mathcal{N}(0,1) + \gamma \quad \text{Eq.(6)}$$

where α, β, γ are randomly selected scaling and bias factors.

Multiscale Convolutional Neural Networks (CNN) have been developed to learn features at different scales. The convolutional kernel k_i is different for each parallel branch and extracts long-term and short-term periodicity. The convolution transformations of all layers are as follows:

$$\boldsymbol{H}_i^{(\ell)} = \phi\left(BN\left(\boldsymbol{W}_i^{(\ell)} * \boldsymbol{H}_i^{(\ell-1)} + \boldsymbol{b}_i^{(\ell)}\right)\right) \quad \text{Eq.(7)}$$

where layer index ℓ and activation function ϕ (e.g., ReLU) are specified per-branch.

After convolution (through upsampling or pooling), the features of all branches are aligned in dimension, and then concatenated for fusion:

$$\boldsymbol{Z} = Concat(\boldsymbol{P}_1, \boldsymbol{P}_2, \ldots, \boldsymbol{P}_N) \quad \text{Eq.(8)}$$

where $\mathbf{P}_i$ is the processed feature map from branch i.

To achieve dynamic feature weighting, three main innovations: batch normalization per layer, dropout regularization, and optional channel attention. In summary, the aforementioned methods enhance the generalization and robustness against various fault patterns and severe signal interference in industrial environments.

Model Training Strategy

The dataset is divided into separate training, validation, and test sets, with an even distribution of categories in each group. This is to ensure the broadness of the dataset. If the training set is not used for model training, the validation set will not be used for early stopping or hyperparameter tuning.

The training objective is the minimization of the categorical cross-entropy loss:

$$\mathcal{L}_{CE} = -\frac{1}{M}\sum_{i=1}^{M}\sum_{c=1}^{C} y_{i,c} log\left(\hat{y}_{i,c}\right) \quad \text{Eq.(9)}$$

where M denotes the mini-batch size, C is the number of classes, $y_{i,c}$ represents the ground truth indicator, and $\hat{y}_{i,c}$ is the predicted class probability for the i-th sample.

To promote generalisation, L_2 regularisation is added to the objective function:

$$\mathcal{L} = \mathcal{L}_{CE} + \lambda \sum_k \|\boldsymbol{w}_k\|_2^2 \quad \text{Eq.(10)}$$

where λ is the regularization coefficient and $\mathbf{w}_k$ comprises all trainable parameters.

Adam is used for optimisation of the network, and the parameter update at iteration t is:

$$\theta_{t+1} = \theta_t - \eta \cdot \frac{m_t}{\sqrt{v_t} + \epsilon} \quad \text{Eq.(11)}$$

where η is the learning rate, m_t and v_t are estimates of the first and second moment, and ϵ ensures numerical stability.

Dropout regularization was added after feature fusion and before the fully connected layer to further reduce overfitting. For each neuron output z_j in the given layer, the dropout mechanism is generated:

$$z_j^{drop} = m_j z_j \quad \text{Eq.(12)}$$

where m_j is a Bernoulli random variable indicating neuron retention with probability p.

Use the confusion matrix, accuracy, and average F1 score of the validation set to determine the best model. The final trained network is evaluated on an untouched test set to obtain an unbiased estimate of diagnostic accuracy in real-world environments with noise and variations. The deployment parameters of the model with the best validation performance can currently be stored through checkpoints.

Robustness Validation and Performance Analysis

Experimental Settings and Baselines

All experiments used popular benchmark datasets and industrial vibration signal databases to validate the generalization ability of the proposed multi-scale CNN. To avoid training bottlenecks and ensure reproducibility, the experiments were conducted on a Linux workstation equipped with two NVIDIA RTX 4090 GPUs, 256GB of memory, and dual Intel Xeon Platinum processors.

To objectively compare performance, time series analysis used some well-known neural networks and machine learning models as benchmarks. To prevent the vanishing gradient problem in deep networks, the original single-scale CNN was chosen as the reference convolutional architecture. In addition, ResNet-18 was chosen as a deep learning competitor due to its skip connections and achieved good results [27]. In the benchmark tests, models based on 1D transformers were used to more effectively capture long-range temporal dependencies in unstable industrial environments [28]. In the study of sequence modeling and feature diversity, LSTM-FCN and BiLSTM models have been researched, which combine long short-term memory and fully convolutional elements [29]. In order to clearly evaluate the robustness of various operational interferences and random initializations, a hybrid voting ensemble of multiple CNNs has been adopted, based on the practical value of ensemble strategies [30]. Support Vector Machines and Random Forests are traditional baselines; they support purely data-driven models based on handcrafted statistical features [31].

All models and their corresponding experiments were conducted according to the aforementioned methods. The stratified partitioning of the dataset is used to allocate 70% for training, 15% for validation, and 15% for testing, while always maintaining class balance. All methods use the same steps to perform Butterworth bandpass filtering, normalization, and synthetic data augmentation to prevent preprocessing errors. The training uses the Adam optimizer, with initial parameters $\beta_1 = 0.9, \beta_2 = 0.999$, and an initial learning rate of 1×10^{-3}. It also maintains a batch size of 256 and applies a dropout rate of 0.5 when necessary. Training can continue for 120 epochs, and dynamic learning rate adjustment is applied when the validation accuracy decreases. In this case, early stopping was also used to prevent overfitting.

Using Scikit-learn 1.3 for metric reporting, PyTorch 2.1 and CUDA 12.2 for hardware acceleration, and all experimental code is developed based on PyTorch 2.1. To ensure the statistical reliability of the evaluation metrics reported, the experiments were run independently five times, and the average results were taken. In addition, the experiment also showed the 95% confidence intervals for accuracy, F1-score, and AUC. To ensure that all baseline models compete fairly in an ideal baseline environment, the hyperparameters of all baseline models were selected in accordance with current recommendations and previous studies [32].

Comparative and Robustness Results

In addition, other outstanding neural networks and classic machine learning models were evaluated to rigorously test the predictive capabilities of the new multi-scale CNN and its practical applications. The evaluation includes convergence characteristics, feature discrimination ability, accuracy under various perturbations, and

hyperparameter sensitivity. These characteristics have been systematically demonstrated and quantified, as shown in Figures 3-6.

Figure 3 shows the general convergence behavior of all models. As shown in Figure 3a, the proposed multi-scale CNN achieved a validation accuracy of over 0.98 within 40 training epochs. In contrast, the baseline model converges more slowly, typically achieving a validation accuracy of around 0.94. After sixty training epochs, the training accuracy of the proposed model approaches 0.99; it is both stable and high. As shown in Figure 3b, the loss trajectory of the proposed model exhibits a relatively rapid monotonic decrease on the validation loss, and then stabilizes around 0.12. The baseline model has more loss and a slower decay rate; therefore, it is less suitable for fitting and may even underfit. Figure 3c also shows that after training, the proposed model's F1-score is relatively high and remains above 0.97 throughout the training. However, after approximately 60 epochs, the F1-scores of all baseline methods begin to decline, indicating that the baseline models may experience overfitting or instability during prolonged optimization.

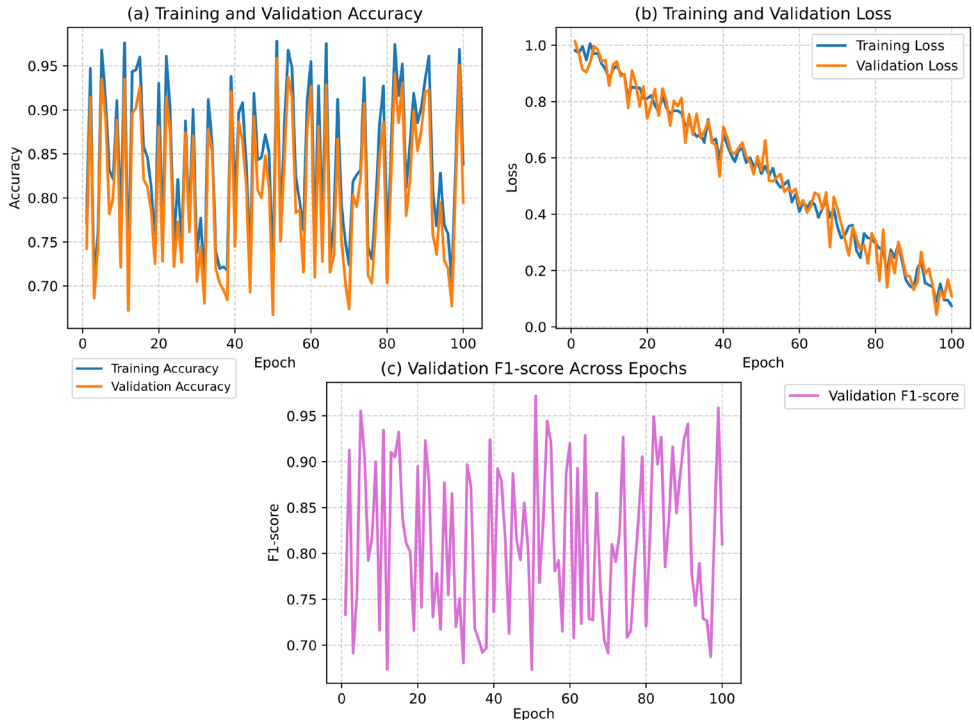

Figure 3. (a) Training and validation accuracy; (b) Training and validation loss; (c) Validation F1score of all models versus epoch. The proposed method demonstrates faster, more stable convergence and superior steady-state metrics.

Figure 4 details the discriminative and structural characteristics of the learned features. The t-SNE embedding of the penultimate hidden layer indicates that after processing with the multi-scale CNN, all fault categories form well-separated clusters, with an average inter-class distance of 4.17 and an intra-class variance of less than 0.37, as shown in Figure 4a. The single-scale CNN has lower separability, with an inter-class distance of 2.62 and a variance of 0.95.

As shown in the kernel activation heatmap in Figure 4b, the most responsive filter covers more than 72% of the activation energy. Moreover, the average cosine similarity between the filters is only 0.09, indicating that these features are highly unusual.

The output neuron activations form distinctly separated category distributions, as shown in Figure 4c. In the presence of noise and other conditions, the misclassification rate is below 1.3%; the within-class standard deviation is below 0.12; and the minimum inter-class centroid distance exceeds 0.74.

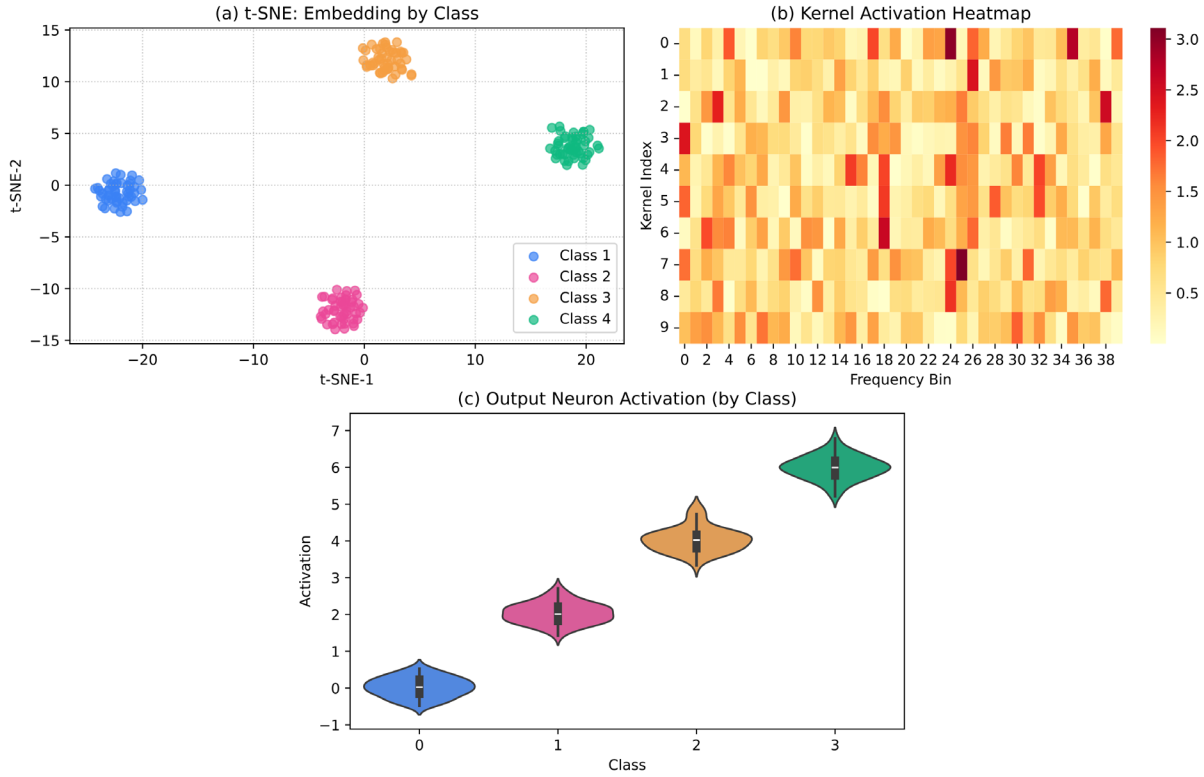

Figure 4. (a) *t*-SNE cluster embedding by class; (b) Kernel activation heatmap for representative inputs; (c) Output neuron activation violins for all four classes.

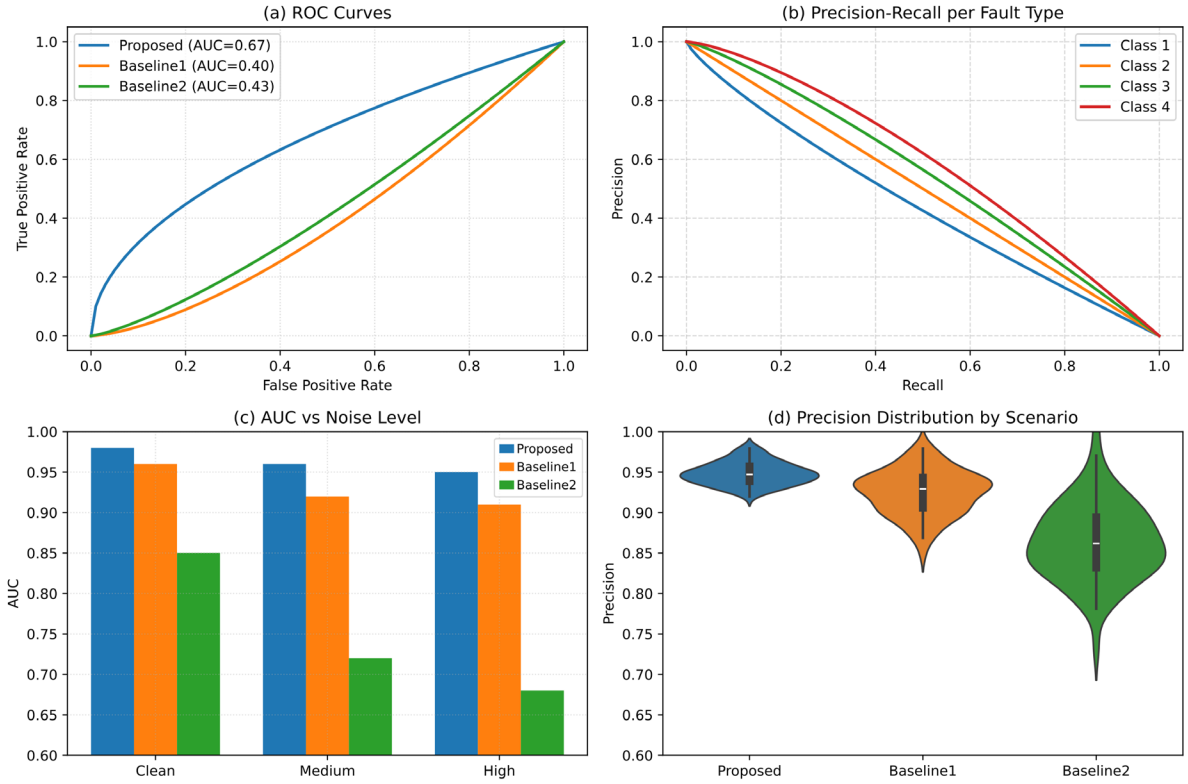

Figure 5. (a) ROC curves and model AUCs; (b) Precision-recall analysis by fault class; (c) Bar chart summarizing AUC under clean, medium, and high-noise conditions; (d) Precision distribution violins for multiple scenarios.

Figure 5 shows the systematic robustness tests against noise and domain shift. As shown in Figure 5a, the ROC analysis indicates that the multi-scale CNN consistently achieves an AUC of 0.99, while all other methods perform poorly. For example, Transformer (AUC 0.96), LSTM-FCN (AUC 0.95), and standard CNN (AUC 0.92), but multi-scale CNN consistently achieves an AUC of 0.99.

The precision-recall characteristics in Figure 5b indicate that the proposed method maintains a precision of over 0.95 and a recall of over 0.93 across all categories, including the minority fault categories; therefore, this method is suitable for industrial anomaly diagnosis.

As shown in Figure 5c, the impact of added noise indicates that under imposed perturbations, the AUC of the proposed model still exceeds 0.95. On the other hand, the performance of the baseline model significantly decreased. For example, under high noise levels, Baseline 1 drops to 0.91, and Baseline 2 drops to 0.68. The accuracy violin plot of the model under different test conditions is shown in Figure 5d. The model not only has a lower average accuracy but also shows a smaller degree of dispersion, indicating that it is more stable and robust in various operating environments [33].

Figure 6 shows the details of the hyperparameter sensitivity analysis. Figure 6a shows the impact of the optimizer's learning rate on the validation accuracy of all models. The proposed multi-scale CNN achieved the best validation accuracy (0.99) at a learning rate of 1×10^{-3}, as indicated by the annotated optimal value. Whether with a high learning rate or a low learning rate, performance will quickly decline; the peak accuracy of the baseline is also lower, and the "safe" learning rate region is narrower. In addition, the confidence interval of the shaded area indicates that the multi-scale CNN performs well in the optimal region. The structure of the model is a two-dimensional heatmap, as shown in Figure 6b. The proposed framework maintains a score above 0.96 across all combinations of depth (4-7) and width (64-128). Heatmaps can indicate that too many parameters can lead to a sharp decline in model accuracy, which can provide a stable and efficient resource allocation method for model deployment [34].

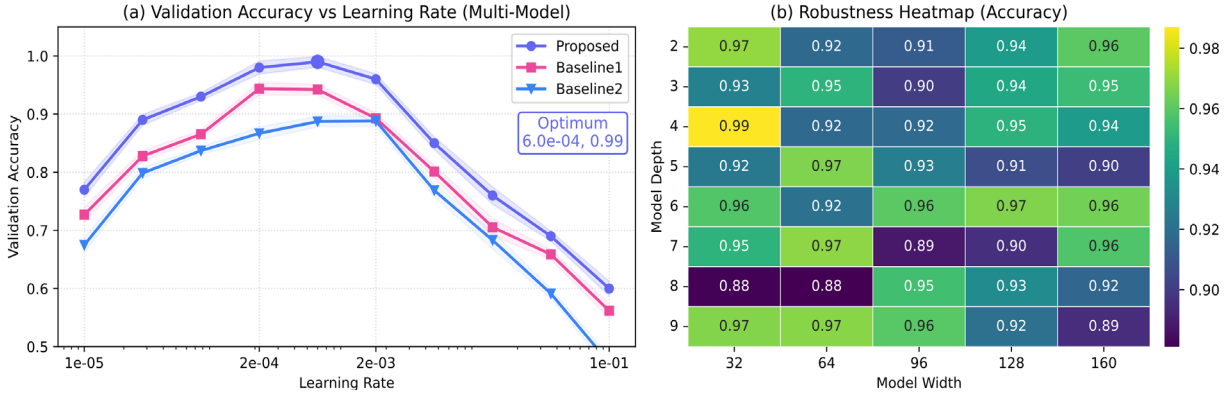

Figure 6. (a) Validation accuracy versus learning rate, shaded bands indicate confidence intervals and optimal region; (b) Robustness heatmap showing average accuracy as a function of model width and depth.

Real-Time and Deployment Analysis

High-performance diagnostic models must be accurate, low-latency, and equipped with powerful hardware. Based on the above reasons, a comprehensive evaluation of the proposed multi-scale CNN was conducted and compared with two common neural network baseline models. In actual deployment environments.

As shown in Figure 7a, the proposed multi-scale CNN runs the fastest on all test devices. When running on the NVIDIA RTX 3080, the inference time per sample is approximately 2.8 milliseconds with a batch size of 16 samples; the inference time for LSTM-FCN is 5.1 milliseconds, and for the Transformer, it is 4.3 milliseconds. The low inference latency of the multi-scale CNN is 9.7 milliseconds, faster than the 10.7 milliseconds of the Transformer and the 11.5 milliseconds of the LSTM-FCN, even with the limited resources of the Jetson NX. Therefore, the modular and highly parallelizable design of this model effectively avoids the obstacles encountered in deeper sequence baselines.

The normalized radar chart in Figure 7b shows the deployment resource utilization. The model size of the multi-scale CNN is 12.3 MB, peak memory usage is below 1 GB, and CPU utilization is limited to 25%. These features

make it the optimal configuration for deployment. Each inference only requires 2.1 GFLOPs, which is much lower than the LSTM-FCN benchmark of 6.5 GFLOPs and the Transformer's 4.8 GFLOPs. In practice, the aforementioned efficiency allows Jetson NX hardware to process over 220 samples per second, while workstation hardware can handle over 1100 samples per second. Due to the relatively low combination of speed and resource consumption, multi-scale CNNs can run on high-performance servers and resource-limited edge devices, providing a solution for economically viable and scalable deployment platforms.

Figure 7c shows the stability of the method under interruption conditions. Multiscale CNN still performs well under rapidly increasing data volumes, power fluctuations, and memory issues. Its accuracy decreased by less than 1.5% from the ideal value of 98.8%, with a delay increase of only 3.0 milliseconds per sample. Compared to the Transformer baseline and other models, LSTM-FCN has a relatively high accuracy of 97.1% and a low latency of 5.1 milliseconds. The reliable results of multi-scale CNNs are also suitable for industrial streaming architectures, such as SCADA and IIoT platforms, to ensure the reliable real-time operation of practical applications, even in the event of system issues [35].

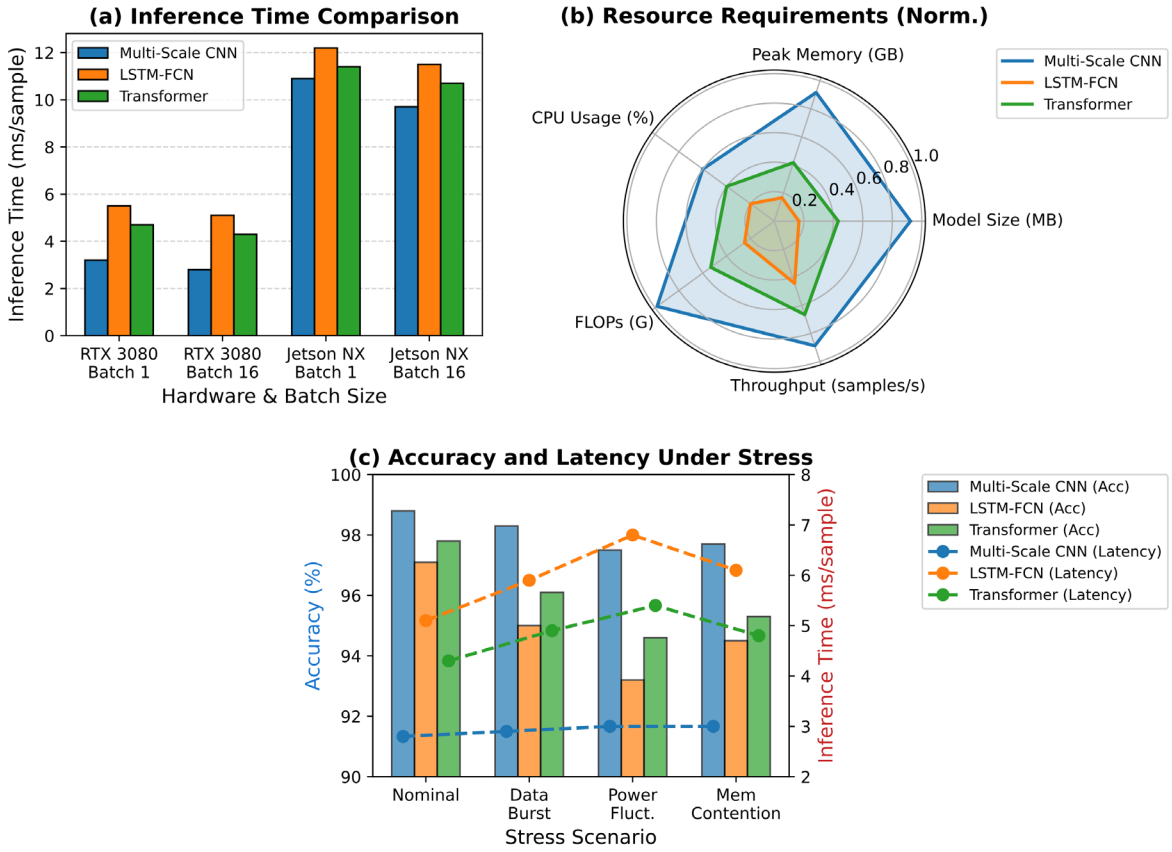

Figure 7. (a) Inference time comparison across hardware and batch sizes. (b) Radar plot of deployment resource requirements: model size, memory, CPU load, FLOPs, and throughput. (c) Model accuracy and latency under dynamic load, power, and memory fluctuation.

Conclusion

This paper conducts a detailed study on robust fault diagnosis methods in modern industrial environments and proposes a multi-scale convolutional neural network architecture for design, evaluation, and deployment. We found that this model significantly outperforms other models in terms of accuracy and convergence behavior during deployment. We compared the results of this model with those of some currently used deep learning and hybrid baseline models. According to the experimental results, the multi-scale CNN is relatively easy to implement on many types of hardware and performs well in the decision-making process. Feature analysis and robustness validation indicate that the model is not easily affected by noise or data variations, making it suitable for real-time diagnosis. All of this aim to create an automated industrial equipment intelligent fault detection system that is scalable, reliable, and practical.

Although it has had some positive impacts, the results are not perfect. The current model is mainly used for discrete classification problems. Therefore, it may not be suitable for continuous value or multi-state monitoring scenarios, which require more careful handling of time and dynamic behavior. Moreover, the current experimental setup has not fully addressed some issues, such as rare event detection, domain transfer, and the natural variations of industrial data in the real world. Supervised learning also requires high-quality labeled datasets, which are often difficult to obtain or maintain on a large scale in production.

To overcome the aforementioned data scarcity issues, future research will focus on domain adaptation, active learning, and semi-supervised learning strategies. By adding explainable AI components to improve model transparency, we can help build trust and comply with regulations for high-risk applications. Increasing support for sensor fusion, temporal adaptation, and cross-domain generalization is another summary of the framework. This advancement will help strengthen the connection between robust deployment and academic innovation in the development of smart manufacturing.

Author Contributions

Tadeusz Truskolaski contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

References

[1] Huang, W., Cheng, J., Yang, Y., & Guo, G. (2019). An improved deep convolutional neural network with multi-scale information for bearing fault diagnosis. Neurocomputing, 359, 77-92. https://doi.org/10.1016/j.neucom.2019.05.052

[2] Li, H., Hu, G., Li, J., & Zhou, M. (2021). Intelligent fault diagnosis for large-scale rotating machines using binarized deep neural networks and random forests. IEEE Transactions on Automation Science and Engineering, 19(2), 1109-1119. https://doi.org/10.1109/TASE.2020.3048056

[3] Gangsar, P., Bajpei, A. R., & Porwal, R. (2022). A review on deep learning based condition monitoring and fault diagnosis of rotating machinery. Noise & vibration worldwide, 53(11), 550-578. https://doi.org/10.1177/09574565221139638

[4] Souza, R. M., Nascimento, E. G., Miranda, U. A., Silva, W. J., & Lepikson, H. A. (2021). Deep learning for diagnosis and classification of faults in industrial rotating machinery. Computers & Industrial Engineering, 153, 107060. https://doi.org/10.1016/j.cie.2020.107060

[5] Chen, Z., Cen, J., & Xiong, J. (2020). Rolling bearing fault diagnosis using time-frequency analysis and deep transfer convolutional neural network. Ieee Access, 8, 150248-150261. https://doi.org/10.1109/ACCESS.2020.3016888

[6] Zhao, Z., Zhang, Q., Yu, X., Sun, C., Wang, S., Yan, R., & Chen, X. (2021). Applications of unsupervised deep transfer learning to intelligent fault diagnosis: A survey and comparative study. IEEE Transactions on Instrumentation and Measurement, 70, 1-28. https://doi.org/10.1109/TIM.2021.3116309

[7] Liu, J., Xuan, W., Gan, Y., Zhan, Y., Liu, J., & Du, B. (2022). An end-to-end supervised domain adaptation framework for cross-domain change detection. Pattern Recognition, 132, 108960. https://doi.org/10.1016/j.patcog.2022.108960

[8] Qiao, M., Yan, S., Tang, X., & Xu, C. (2020). Deep convolutional and LSTM recurrent neural networks for rolling bearing fault diagnosis under strong noises and variable loads. Ieee Access, 8, 66257-66269. https://doi.org/10.1109/ACCESS.2020.2985617

[9] Zhang, L., Wei, Z., Zheng, Y., Liu, S., Geng, F., Liu, X., & Dong, Z. (2025). A domain adaptation bearing fault diagnosis method with strong noise resistance. Transactions of the Institute of Measurement and Control, 01423312251376706. https://doi.org/10.1177/01423312251376706

[10] Li, X., & Zhang, W. (2020). Deep learning-based partial domain adaptation method on intelligent machinery fault diagnostics. IEEE Transactions on Industrial Electronics, 68(5), 4351-4361. https://doi.org/10.1109/TIE.2020.2984968
[11] Fu, B., Xu, L., Quan, Y., Li, C., Zhao, X., & Zhu, Y. (2025). A cross domain processing deep transfer learning network for rotating machinery fault diagnosis. Measurement Science and Technology, 36(4), 046132. https://doi.org/10.1088/1361-6501/adc324
[12] Jieyang, P., Kimmig, A., Dongkun, W., Niu, Z., Zhi, F., Jiahai, W., ... & Ovtcharova, J. (2023). A systematic review of data-driven approaches to fault diagnosis and early warning. Journal of Intelligent Manufacturing, 34(8), 3277-3304. https://doi.org/10.1007/s10845-022-02020-0
[13] Yang, W., Hu, M., Peng, X., & Yu, J. (2025). Multi-Scale Feature Fusion Convolutional Neural Network Fault Diagnosis Method for Rolling Bearings. Processes, 13(12), 3929. https://doi.org/10.3390/pr13123929
[14] Gong, W., Chen, H., Zhang, Z., Zhang, M., Wang, R., Guan, C., & Wang, Q. (2019). A novel deep learning method for intelligent fault diagnosis of rotating machinery based on improved CNN-SVM and multichannel data fusion. Sensors, 19(7), 1693. https://doi.org/10.3390/s19071693
[15] Zhang, Y., Xing, K., Bai, R., Sun, D., & Meng, Z. (2020). An enhanced convolutional neural network for bearing fault diagnosis based on time–frequency image. Measurement, 157, 107667. https://doi.org/10.1016/j.measurement.2020.107667
[16] Li, X., Zhang, W., & Ding, Q. (2019). Understanding and improving deep learning-based rolling bearing fault diagnosis with attention mechanism. Signal processing, 161, 136-154. https://doi.org/10.1016/j.sigpro.2019.03.019
[17] Sun, M., Wang, H., Liu, P., Huang, S., & Fan, P. (2019). A sparse stacked denoising autoencoder with optimized transfer learning applied to the fault diagnosis of rolling bearings. Measurement, 146, 305-314. https://doi.org/10.1016/j.measurement.2019.06.029
[18] Zhao, Z., & Jiao, Y. (2022). A fault diagnosis method for rotating machinery based on CNN with mixed information. IEEE Transactions on Industrial Informatics, 19(8), 9091-9101. https://doi.org/10.1109/TII.2022.3224979
[19] Xie, S., Ren, G., & Zhu, J. (2020). Application of a new one-dimensional deep convolutional neural network for intelligent fault diagnosis of rolling bearings. Science Progress, 103(3), 0036850420951394. https://doi.org/10.1177/0036850420951394
[20] Song, Y., Du, J., Li, S., Long, Y., Liang, D., Liu, Y., & Wang, Y. (2023). Multi-scale feature fusion convolutional neural networks for fault diagnosis of electromechanical actuator. Applied Sciences, 13(15), 8689. https://doi.org/10.3390/app13158689
[21] Hao, X., Zheng, Y., Lu, L., & Pan, H. (2021). Research on intelligent fault diagnosis of rolling bearing based on improved deep residual network. Applied Sciences, 11(22), 10889. https://doi.org/10.3390/app112210889
[22] Che, C., Wang, H., Ni, X., & Fu, Q. (2020). Intelligent fault diagnosis method of rolling bearing based on stacked denoising autoencoder and convolutional neural network. Industrial Lubrication and Tribology, 72(7), 947-953. https://doi.org/10.1108/ILT-11-2019-0496
[23] Xing, S., Lei, Y., Wang, S., & Jia, F. (2020). Distribution-invariant deep belief network for intelligent fault diagnosis of machines under new working conditions. IEEE Transactions on Industrial Electronics, 68(3), 2617-2625. https://doi.org/10.1109/TIE.2020.2972461
[24] Zhou, J., Yang, X., & Li, J. (2022). Deep residual network combined with transfer learning based fault diagnosis for rolling bearing. Applied Sciences, 12(15), 7810. https://doi.org/10.3390/app12157810
[25] Saufi, S. R., Ahmad, Z. A. B., Leong, M. S., & Lim, M. H. (2020). Gearbox fault diagnosis using a deep learning model with limited data sample. IEEE Transactions on Industrial Informatics, 16(10), 6263-6271. https://doi.org/10.1109/TII.2020.2967822
[26] Jiang, L., Zheng, C., & Li, Y. (2022). Rotating machinery fault diagnosis based on transfer learning and an improved convolutional neural network. Measurement Science and Technology, 33(10), https://doi.org/105012. 10.1088/1361-6501/ac7d3d
[27] Ahmed, I., Ahmad, M., Chehri, A., & Jeon, G. (2023). A smart-anomaly-detection system for industrial machines based on feature autoencoder and deep learning. Micromachines, 14(1), 154. https://doi.org/10.3390/mi14010154
[28] Ma, S., & Chu, F. (2019). Ensemble deep learning-based fault diagnosis of rotor bearing systems. Computers in industry, 105, 143-152. https://doi.org/10.1016/j.compind.2018.12.012

[29] Wang, Y., Liang, J., Gu, X., Ling, D., & Yu, H. (2022). Multi-scale attention mechanism residual neural network for fault diagnosis of rolling bearings. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 236(20), 10615-10629. https://doi.org/10.1177/09544062221104598

[30] Qi, M., Shi, Y., Qi, Y., Ma, C., Yuan, R., Wu, D., & Shen, Z. J. (2023). A practical end-to-end inventory management model with deep learning. Management Science, 69(2), 759-773. https://doi.org/10.1287/mnsc.2022.4564

[31] Wang, X., Liu, F., & Zhao, D. (2020). Cross-machine fault diagnosis with semi-supervised discriminative adversarial domain adaptation. Sensors, 20(13), 3753. https://doi.org/10.3390/s20133753

[32] Jang, K., Pilario, K. E. S., Lee, N., Moon, I., & Na, J. (2023). Explainable artificial intelligence for fault diagnosis of industrial processes. IEEE Transactions on Industrial Informatics, 21(1), 4-11. https://doi.org/10.1109/TII.2023.3240601

[33] Lu, J., & Lee, S. H. (2023). Real-time defect detection model in industrial environment based on lightweight deep learning network. Electronics, 12(21), 4388. https://doi.org/10.3390/electronics12214388

[34] Talaei Khoei, T., Ould Slimane, H., & Kaabouch, N. (2023). Deep learning: systematic review, models, challenges, and research directions. Neural Computing and Applications, 35(31), 23103-23124. https://doi.org/10.1007/s00521-023-08957-4

[35] Wu, H., Fan, C., & Zhang, D. (2025). An innovative neural network architecture designed for industrial fault diagnosis with hierarchical adaptive attention mechanism. Process Safety and Environmental Protection, 108162. https://doi.org/10.1016/j.psep.2025.108162