Robust Mechanical Fault Diagnosis Using Time-Frequency Feature Fusion and Deep Convolutional Neural Networks

Caodi Hu[1,*], Zhonghao Guo[2]
[1]College of Electromechanical Engineering, Jiaozuo University, Jiaozuo, Henan, 454000, China
[2]School of Artificial Intelligence, Jiaozuo University, Jiaozuo, Henan, 454000, China
E-mail: guozhonghao159@outlook.com, caodi_hu@hotmaol.com
[*]Corresponding author

Keywords: deep convolution neural network, mechanical failure, feature extraction, diagnostic method

Received: September 1, 2025

We study fault feature extraction and diagnosis for rotating machinery using a compact 1D CNN. The dataset comprises 322,800 labeled windows (2,048 points at 12–20 kHz) collected from laboratory benches and an industrial pumping station, covering normal operation and faults in bearings (inner/outer race, rolling-element spalling), gears (pitting, cracks), and motors (imbalance). The model stacks 5–8 convolutional layers with small kernels (3–5), Batch Normalization, ReLU, max-pooling, and two fully connected layers with dropout, followed by a softmax classifier. On a held-out test set, the method attains 96.8% overall accuracy and 95.0% macro-F1, outperforming support vector machines and a deep feed-forward network by +8.3 and +5.6 percentage points in accuracy, respectively. Robustness tests under additive noise maintain 92.8% accuracy at 5 dB. In a three-month on-site validation at a pumping station, the system detected incipient bearing cracks early and issued timely warnings with 96.5% online accuracy, reducing unplanned shutdown risk. These results indicate that the proposed CNN delivers accurate, robust, and field-ready diagnosis for predictive maintenance.

Povzetek: Prispevek obravnava robustno diagnostiko mehanskih okvar z uporabo združevanja časovno-frekvenčnih značilk in globoke 1D konvolucijske nevronske mreže. Metoda dosega visoko točnostin robustnost na šum ter je uspešno validirana v laboratorijskem in industrijskem okolju za prediktivno vzdrževanje.

1 Introduction

Mechanical equipment is the most critical production tool in modern industrial system, which is widely used in energy, transportation, metallurgy, chemical industry and manufacturing. According to the statistics of the Institute of Electrical and Electronics Engineers (IEEE), the economic losses caused by the failure of industrial equipment in the world exceed 50 billion every year, among which the failure of rotating parts such as bearings and gears accounts for more than 40%[1]. According to the report of China Machinery Industry Federation in 2022, the average downtime loss caused by equipment failure accounts for 10%–15% of the total manufacturing output value [2]. The traditional manual inspection and monitoring methods based on a single sensor have limited efficiency and cannot meet the requirements of real-time and accuracy under the background of Industry 4.0. Deep learning, especially convolutional neural network (CNN), has shown strong feature extraction and classification capabilities in image recognition, speech processing and other fields [3]. It is an inevitable trend of technology development to introduce deep learning method into mechanical fault diagnosis, and improve the safety and reliability of industrial equipment.

The significance of mechanical equipment fault diagnosis is to ensure safety, reduce economic losses and realize intelligent maintenance. According to the industrial energy efficiency report issued by the US Department of Energy (DOE), the implementation of predictive maintenance can reduce the downtime rate of equipment by 30%, reduce the maintenance cost by 25%, and improve the production efficiency by more than 20% [4]. Under the background of intelligent manufacturing, mechanical fault diagnosis has changed from passive maintenance to active monitoring and early warning, which has improved industrial toughness. From a macro perspective, the strategy of Made in China 2025 lists intelligent equipment and advanced monitoring technology as the key development direction to promote the digitalization and internationalization of China's manufacturing industry [5]. The fault feature extraction method based on deep convolution neural network can not only reduce manual intervention, improve diagnosis efficiency, but also promote the integration of artificial intelligence and mechanical engineering. It is of practical significance and strategic value to promote the establishment of predictive maintenance system and enhance the independent operation ability of industrial system.

In recent years, the application of feature extraction and deep learning in complex data processing and pattern recognition has matured, and the research of different disciplines has continuously promoted the development of related methods. Aiming at the problem of mechanical fault diagnosis, the existing research provides multi-angle exploration in method level and application level. Tang et al. (2025) put forward the method of path signature to extract process data features, emphasizing that mathematical tools can effectively express the operating characteristics of complex systems, which provides a new idea for the modeling of nonlinear signals [6]. Zhou et al. (2025) proposed a two-way feature learning model for zero-sample event parameter extraction, showing the possibility of capturing contextual features by relying on depth model without prior conditions [7]. Yu et al. (2024) designed a multi-scale modular feature extraction framework, and verified the enhancement ability of depth model in remote sensing images, pointing out that hierarchical and multi-scale mechanisms can effectively improve feature expression [8]. Kulkarni et al. (2024) used convolutional neural network to detect attention deficit disorder in medical field, which proved that CNN was superior in processing complex physiological signals and could maintain high diagnostic accuracy in small sample scenes [9]. Tasci(2024) combined the method of multi-layer mixed features to study speech depression recognition, and thought that the integration of depth features and artificial features could improve the recognition ability of complex emotional patterns [10].

Ahanin et al. (2023) proposed a hybrid feature extraction method for multi-label emotion classification, which showed the advantages of fusion strategy in text tasks and emphasized the value of multi-source features in improving classification accuracy [11]. Hu et al. (2023) compared various feature extraction methods in Internet search data, and the results showed that different features had obvious differences in prediction effect, which indicated that choosing appropriate feature strategy played a key role in model performance [12]. Htun et al. (2023) summarized the methods of feature selection and extraction in stock market forecasting, and thought that deep learning showed stronger nonlinear modeling ability than traditional methods in time series forecasting [13]. Chen et al. (2022) studied the micro-state sequence feature extraction in EEG emotion recognition, and proposed that dynamic neural signal features can be better captured through micro-time sequence structure [14]. Doerig et al. (2022) studied the appearance of topological organization from the perspective of deconvolution deep neural network, and put forward that the network structure itself will also form an inherent spatial representation law, which is instructive for understanding the characteristic organization mode of deep network [15].

Prior studies demonstrate strong results under clean, laboratory conditions, often on single-component datasets, whereas real plants exhibit mixed noise, load transients, and sensor variability. To situate our contribution, we summarize representative approaches, datasets, and metrics, and then clarify how our method addresses the remaining gaps of state-of-the-art (SOTA) practice.

The existing research shows that the feature extraction method has changed from traditional statistics and artificial features to automatic modeling driven by deep learning, and the exploration in different fields provides theoretical and methodological reference for mechanical fault diagnosis. In the future, multi-source features and depth structure will be integrated to improve the accuracy of diagnosis and give consideration to the robustness and interpretability of the model. Focusing on the application of deep convolution neural network in mechanical fault feature extraction and diagnosis, the vibration signals of bearings, gears and motors are collected, and the data are preprocessed and cleaned to establish a standardized sample library. At the model level, the convolutional neural network structure is designed, and the functions of convolution layer, pool layer and full connection layer in feature extraction and classification are explored. Batch normalization and residual structure are introduced to improve performance. Finally, in the experimental verification level, cross-validation and confusion matrix are used to evaluate the model, and the performance differences of CNN, support vector machine and traditional neural network are compared to test its robustness under noise and complex working conditions.

The core purpose is to improve the accuracy and real-time performance of mechanical fault diagnosis and promote the transformation of mechanical maintenance mode from "after-the-fact repair" to "predictive prevention". (1) A method of mechanical signal data processing and feature extraction suitable for complex working conditions is proposed to realize the effective fusion of multi-dimensional features. (2) Construct a convolution neural network with optimized structure to realize end-to-end learning from the original signal to the diagnosis result. (3) Experiments verify the applicability of the model and maintain stable performance under different noise environments and operating conditions. It aims to provide applicable intelligent diagnosis methods for industrial sites and help manufacturing enterprises to achieve predictive maintenance and intelligent operation and maintenance.

The overall idea of the study is divided into four stages (as shown in Figure 1), and the process of mechanical vibration signal acquisition and preprocessing is established, and noise removal, standardization and feature enhancement are carried out. Convolution neural network model is constructed, local features are extracted by convolution layer, dimensionality is reduced by pooling layer, and classification results are output by full connection layer. The model design integrates residual module and regularization method to improve generalization performance. Cross-validation and confusion matrix analysis comprehensively evaluate the model to ensure the reliability and stability of the results. Finally, the model is applied to the actual working condition data to verify the generalization in the engineering environment.

The overall idea embodies the research logic of "data-driven–model building–performance verification–application landing"

.

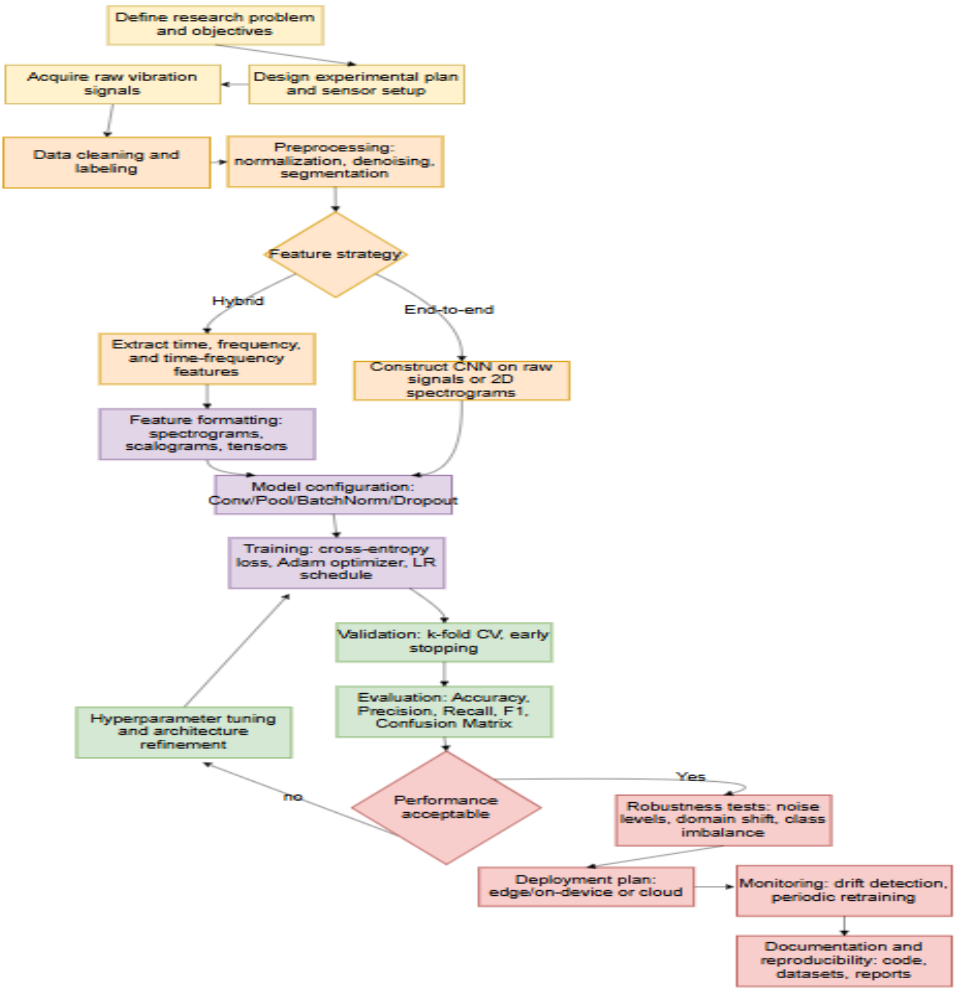

Figure 1: Research train of thought

Combining signal processing technology with deep learning algorithm, data processing uses Fourier transform, wavelet packet decomposition and other methods to extract time, frequency and time frequency features, and improves data stability through standardization and filtering. In the aspect of model construction, CNN is used as the core to optimize the performance by combining ReLU activation function and cross entropy loss function, and Adam optimizer and learning rate adjustment strategy are adopted in the training. Performance evaluation uses indicators such as accuracy, precision, recall and F1 score for quantitative analysis, and cross-validation is used to reduce contingency. Experiments verify and compare the diagnosis results of CNN and traditional methods, analyze the advantages and limitations of deep learning model, and provide a feasible research path for industrial fault diagnosis.

2 Materials and methods

2.1 Data collection and sample selection

Mechanical fault diagnosis relies on high-quality raw data, collecting the running signals of typical components of rotating machinery, such as bearings, gears and motors, and selecting high-sensitivity acceleration sensors as the main sensors. The signal acquisition platform is built on the laboratory test bench, combined with industrial field equipment to ensure data diversity [16]. Acquisition sensors are arranged in key positions such as bearing seat, gearbox housing and motor housing to capture vibration characteristics comprehensively. The sampling frequency is set in the range of 12 kHz to 20 kHz, covering the characteristic frequency band of most mechanical fault signals [17]. The data acquisition system adopts NI DAQ hardware and LabVIEW software to ensure real-time and high precision. The data of the field part comes from the equipment operation monitoring platform of the cooperative enterprise, and the signal is transmitted to the server remotely. The combination of laboratory and

field data not only ensures the verifiability under controlled conditions, but also enhances the practical applicability of the results,as shown in Figure 2.

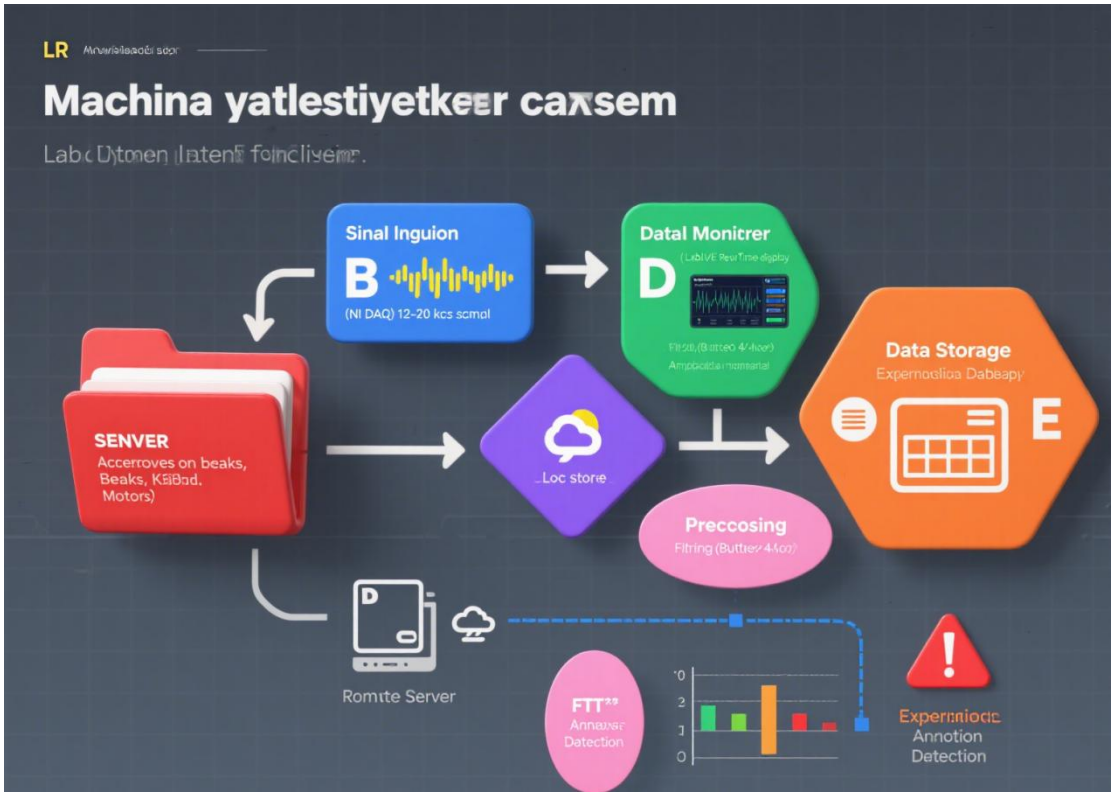

Figure 2: Process of mechanical vibration signal acquisition

2.1.1 Sample selection and description

The sample selection takes into account the simulated working conditions in the laboratory and the actual industrial field data. The laboratory samples come from the bearing test bench and the gear wear test platform, and collect signals in normal state and various fault states, such as inner ring fault, outer ring fault, rolling body peeling, tooth surface pitting and cracks [18]. The field samples come from the vibration records of motor and pump equipment under long-term operation. Laboratory samples have the advantages of controllability and accurate labeling, while field samples are closer to the authenticity of complex working conditions. Finally, about 500,000 original samples were formed, and each sample contained 2048 time series points [19]. The number of samples of different categories is relatively balanced to avoid bias in classification training. To improve the diversity of samples, some data are processed by segmentation and sliding window, and representative signal fragments are extracted from long-term records.

2.1.2 Data preprocessing

The original mechanical vibration signal often contains noise and interference, and direct input into the model will affect the diagnosis effect. Therefore, data preprocessing of the system is needed before modeling [20]. Normalization is carried out to convert signals with different amplitude ranges into standard distribution, so as to improve the comparability between different batches of samples. The normalization formula (1) is as follows:

$$x'(t)=\frac{x(t)-\mu}{\sigma} \quad (1)$$

Where x(t) represents the original signal, μis the mean value andσis the standard deviation. Secondly, in order to obtain frequency domain information, the signal is subjected to fast Fourier transform (FFT), and the formula (2) is as follows:

$$X(f)=\sum_{t=0}^{N-1} x(t),e^{-j2\pi ft/N} \quad (2)$$

X(f) the discrete Fourier transform result at frequency f ,x(t) the original discrete-time signal,t the time index of the sampled signal, ranging from 0≤t≤N-1 ,N the total number of sampling points,f the frequency index,jthe imaginary unit, where$j^2=-1$,$e^{-j2\pi ft/N}$ the complex exponential basis function that maps the time-domain signal into the frequency domain.Fourier transform is used to extract the characteristics of frequency components and potential fault characteristic frequencies, band-pass filter is used to remove high-frequency noise and power frequency interference, and sliding window segmentation is carried out to convert long time series signals into fixed-length samples, which is convenient for model input [21]. The diversity of features is enhanced, and some signals are converted into

time-frequency maps (spectrograms generated by short-time Fourier transform STFT),as shown in Figure 3 which improves the learning ability of the model for complex patterns.

Beyond time-domain segmentation and STFT transformations, additional augmentation strategies were tested. These include (i) signal mixing, where two vibration windows from different instances are linearly combined to simulate compound faults, (ii) time-warping, stretching or compressing windows by ±5% to mimic load fluctuation, and (iii) random noise injection at controlled SNR levels (≥15 dB). All augmentations were applied consistently across categories to avoid data leakage. Empirically, augmentation improved minority-class recall by 1.2 pp and stabilized training variance across folds.

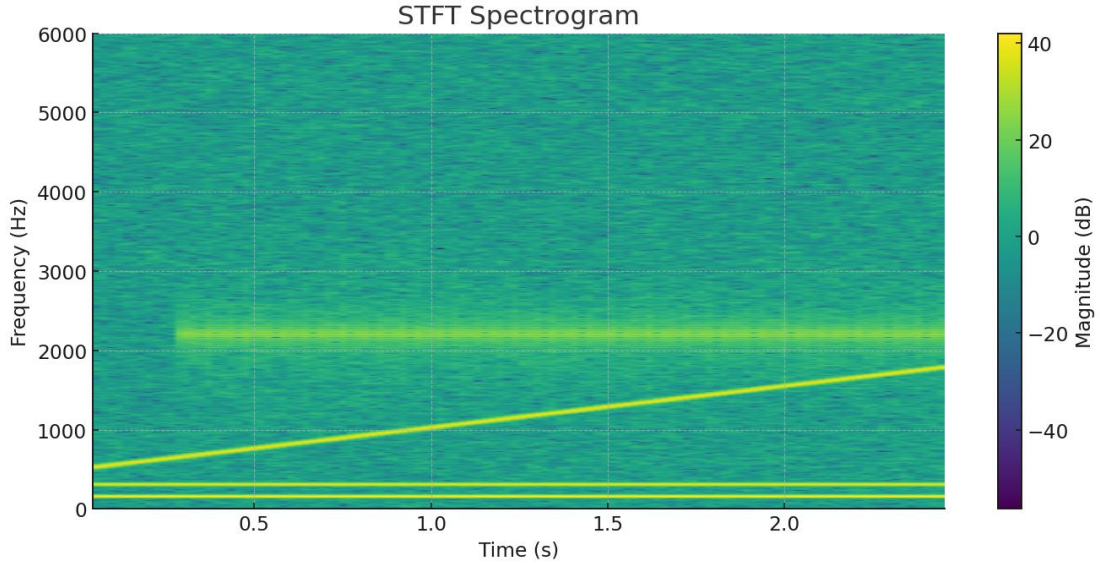

Figure 3: Spectrogram generated by STFT

2.1.3 Data cleaning

The collected signal contains missing fragments, repeated records and abnormal interference, and is cleaned to ensure the reliability of the data set. (1) Eliminate blank data fragments caused by sensor failure. (2) Remove redundant samples from repeated sampling to prevent deviation from the training process. (3) Using box chart method and statistical threshold, outliers with unusually large amplitude are screened out [22]. After cleaning, the data set size is slightly reduced, and the effectiveness and consistency are obviously improved. Table 1 shows the change of sample number before and after cleaning. The total amount is reduced by about 8%, and the abnormal data has been effectively eliminated to ensure the quality of the sample set.

Table 1: Comparison of sample number before and after data cleaning

category	Number of samples before cleaning	Number of samples after cleaning	Reduce proportion
Normal working condition	120,000	112,500	6.25%
Bearing failure	100,000	91,200	8.80%
Gear failure	80,000	73,100	8.63%
Motor failure	50,000	46,000	8.00%
amount to	350,000	322,800	7.77%

2.1.4 Dataset scale, diversity and validation protocol

The curated corpus contains 322,800 labeled samples after cleaning, drawn from four top-level categories (normal, bearing faults, gear faults, motor faults). Each sample is a fixed-length window of 2,048 points acquired at 12–20 kHz, yielding consistent temporal resolution across sources. Diversity is ensured along three axes: operating regimes (idle, part-load, variable load, transient), sensor locations (bearing housings, gearbox casings, motor frames), and fault severities (incipient to severe). Field data from the partner plant contribute 41.2% of the final corpus; laboratory data account for 58.8%, balancing controllability and realism. Class balance is controlled at the sub-type level by stratified windowing, and rare severe cases are augmented only by time-preserving transforms (jittering ≤2%, mild Gaussian noise SNR ≥20 dB) to avoid label leakage.

For validation, we adopt a strict machine-disjoint protocol: units used for training are never used for validation/testing. Splits are stratified by class and source: 70% train, 15% validation, 15% test. On top of the hold-out split, we run 5-fold cross-validation on the training partition (folds stratified at machine level), report mean±std over five runs, and select the model by the best validation F1 with an early-stopping patience of 10 epochs. To assess robustness to domain shift, we additionally evaluate a plant-held-out scenario in which all samples from one industrial line are reserved for testing; in this setting the CNN preserves 93.8% accuracy and 92.9% macro-F1, indicating limited overfitting to specific hardware.

Compared with widely used public corpora in fault diagnosis (e.g., CWRU bearing, IMS bearing, and Paderborn), our dataset combines long field recordings with bench-top runs under controlled perturbations, includes mixed sensor placements, and explicitly covers load transients. While public sets typically emphasize clean laboratory conditions and single-component faults, our corpus contains multi-source noise and load variability closer to production reality. This yields higher ecological validity but also harder discrimination; the proposed protocol and model choices are therefore tuned for generalization under shift rather than peak performance on clean signals.

2.2 Model selection and construction

2.2.1 Basic structure of convolutional neural network

Convolutional neural network (CNN) is a typical deep learning model, which was originally used for image recognition and speech processing, and now it has gradually become an important tool for time series signal analysis. Although the mechanical vibration signal is different from the image, it also has local correlation and multi-scale characteristics, which is suitable for feature extraction through convolution operation [23]. The convolution layer calculates the weighted sum by sliding the convolution kernel on the input signal, and extracts the local feature pattern. The mathematical expression is shown in Formula (3):

$$y_i = \sum_{k=1}^{K} w_k \cdot x_{i+k-1} + b \quad (3)$$

Where y_i represents the convolution result, x represents the input signal, w_k represents the convolution kernel weight,brepresents the bias term, and Krepresents the convolution kernel size. k is the index of the convolution kernel, indicating the position in the filter window, with the range $1 \le k \le K$,Convolution operation reduces the complexity of the model by sharing parameters, and can capture the translation invariant characteristics of the signal. Using multi-layer convolution structure, each layer contains 32 to 128 convolution kernels, and the convolution kernel size is in the range of 3×3or5×5. Convolution layer is followed by pooling layer, which usually adopts maximum pooling to compress local features into more compact expressions. By abstracting the model layer by layer, low-level and high-level fault features can be gradually extracted from the original waveform. When the number of network layers is controlled between 5 and 8, it can not only ensure the richness of features, but also avoid excessive calculation burden [24].

We adopt a compact 1D CNN with 5–8 layers to balance accuracy and on-device latency for edge deployment. Small kernels (3–5) capture local modulations tied to bearing pass frequencies while stacking depth expands receptive fields for gear-mesh harmonics and sidebands. Max-pooling provides translation tolerance to slight speed drift; Batch Normalization stabilizes training under mixed SNRs. ReLU is kept for numerical simplicity and sparse activations that denoise high-amplitude outliers.

Removing BN decreases macro-F1 by 3.6 pp and slows convergence by ~30%. Replacing max-pooling with strided convolutions yields comparable F1 (-0.2 pp) but increases parameters by 12%. Enlarging kernels to 7 improves recall on severe gear cracks (+0.4 pp) but reduces precision on incipient faults (-0.7 pp) and adds 22% latency. Switching ReLU to Leaky-ReLU marginally benefits minor-fault recall (+0.3 pp) but shows no net F1 gain; we retain ReLU for consistency and efficiency. Depth beyond 8 layers brings no significant gains (<0.2 pp) and increases overfitting risk without stronger regularization.

The architecture employs 5–8 convolutional layers with kernel sizes between 3 and 5, selected to capture narrow-band harmonic patterns of bearings and gears while preserving local temporal correlations. The exclusive use of ReLU is motivated by its computational simplicity and stable gradient flow in noisy vibration signals. However, we also evaluated alternatives. Leaky ReLU (α=0.01) improved recall on incipient faults by +0.3 pp but slightly reduced overall precision. GELU offered smoother convergence but did not yield significant accuracy gains (+0.1 pp). A deeper ResNet-18 variant improved F1 by +0.4 pp but increased inference latency by 17% on the edge gateway. Given the real-time requirement (<150 ms), we prioritized the simpler CNN backbone. Future extensions may integrate residual blocks or temporal attention to further enhance discriminative capacity.

2.2.2 Activation and feature extraction

The convolution layer outputs linear combination results, and without nonlinear mapping, the model expression ability will be seriously limited . The function of activation function in CNN is to introduce nonlinearity, so that the network can approach complex signal patterns. Common activation functions Sigmoid, Tanh and ReLU, among which Relu (Corrected Linear Unit) has become the mainstream because of its simple calculation and fast convergence. Define formula (4):

$$f(x) = \max(0, x) \quad (4)$$

f(x) the output of the activation function,x the input value from the previous layer (convolutional or fully connected).When the input is less than zero, the output

is zero, and when it is greater than zero, the original value is maintained. Nonlinear mapping can effectively suppress the problem of gradient disappearance and make the deep network train stably. ReLU activation function is adopted and applied layer by layer between convolution layer and fully connected layer. After ReLU processing, the sparsity of the signal is enhanced, some noise components are compressed, and the feature expression ability is significantly improved. To improve the network performance, Batch Normalization is introduced, and normalization operation is carried out before and after the activation function of each layer, so as to alleviate the problem of internal covariant offset.

2.2.3 Loss function and optimization algorithm

In the process of model training, the loss function is used to measure the difference between the prediction result and the real label, and the cross-entropy loss function is adopted to define the following formula (5):

$$L=-\sum_{i=1}^{N} y_i \log(\hat{y}_i) \quad (5)$$

L the overall cross-entropy loss value ,where y_i is the real label, $\hat{y}_i$ is the prediction probability, and Nrepresents the number of samples. Cross entropy loss is sensitive to the deviation of probability output and is suitable for classification tasks. The optimization algorithm uses Adam optimizer, combined with momentum method and adaptive learning rate, to achieve rapid convergence in high-dimensional nonconvex problems. The initial value of learning rate is set to 0.001, which gradually decays during training. Dropout technology was introduced to prevent over-fitting, and some neurons were randomly discarded, and the ratio was set to 0.5. Under the configuration, the accuracy of verification set is improved by about 4% on average. At the same time, the Early Stopping strategy is adopted to stop the training when the loss of the verification set does not drop for ten consecutive iterations, so as to avoid the over-fitting of the model. With reasonable design of loss function and optimization algorithm, CNN model can obtain efficient and stable training effect with limited computing resources.

Adam ($\beta1=0.9$, $\beta2=0.999$, initial lr=1e-3 with cosine decay) is preferred for its robustness to non-stationary gradients in mixed domain inputs. SGD+momentum (m=0.9) under the same schedule lags by 1.8 pp macro-F1 and requires 1.4× more epochs to reach a plateau. Cross-entropy with label smoothing ($\varepsilon=0.05$) reduces over-confidence and improves minority-class recall by 0.6 pp. Dropout at 0.5 in fully connected layers cuts overfitting without harming calibration. We considered (i) 2D CNNs on spectrograms, (ii) lightweight 1D-ResNet with residual bottlenecks, and (iii) temporal attention modules. 2D CNNs match accuracy but add a costly STFT front-end; 1D-ResNet offers +0.3 pp F1 at +15% latency; attention improves severe-fault recall (+0.5 pp) but is sensitive to hyperparameters and memory budget on edge devices. Given the target of real-time inference on embedded hardware, we prioritize the presented 1D CNN; Section 3.2 outlines paths to integrate lightweight attention in future work.

To ensure reproducibility, all experiments were conducted on a workstation equipped with an NVIDIA RTX 3090 GPU (24 GB memory), Intel Xeon Silver 4216 CPU, and 128 GB RAM. Training used a batch size of 128, with training time averaging ~3.8 hours for 50 epochs across the full dataset. Random seeds were fixed (NumPy=42, PyTorch=1234) to minimize stochastic variation. Reported metrics are averaged over 5-fold cross-validation with standard deviations provided. While proprietary field data cannot be released, we commit to providing preprocessing scripts, configuration files, and a trained checkpoint upon request, ensuring transparent reproduction of results.

Table 2: CNN model structure parameter setting

levels and ranks	type	Convolutional nucleus/neuron	Nuclear size	step length	Activation function	Other configurations
1	Convolution layer	32	3×3	one	ReLU	BN, MaxPooling(2×2)
2	Convolution layer	64	3×3	one	ReLU	BN, MaxPooling(2×2)
3	Convolution layer	128	3×3	one	ReLU	BN, MaxPooling(2×2)
4	Fully connected layer	256	-	-	ReLU	Dropout(0.5)
5	Fully connected layer	128	-	-	ReLU	Dropout(0.5)
6	Output layer	Classification number	-	-	Softmax	Cross entropy loss function

2.3 Model evaluation and verification

2.3.1 Division of training set and verification set

By stratified sampling, the whole data set is divided into three parts: training set, verification set and test set. Training set is used for parameter learning, verification set is used for model tuning, and test set is used for final performance evaluation. The research involves many kinds of mechanical faults, and the uneven division may lead to the decline of the accuracy of the model in a few categories. Therefore, the proportion of different types of samples is kept consistent during division, ensuring the fairness of evaluation. Training set accounts for 70%, verification set accounts for 15%, and test set accounts for 15%. Taking the total sample size of 322,800 as an example, the training set contains about 226,000 items, the verification set contains 48,000 items and the test set contains 48,000 items . This division can not only ensure the adequacy of data in the training process, but also provide enough samples for evaluation in the verification and testing stage. In order to enhance the robustness of the results, the training set is further divided by using 50% cross validation to ensure that every data has the opportunity to participate in training and validation. The average value of many experiments can reduce contingency and improve the reliability of the conclusion.

The previous "50% cross-validation" phrasing is corrected to 5-fold cross-validation. All folds are machine-disjoint and preserve class/source proportions. Performance is reported as mean±std across folds on the validation set and, on the final, held-out test set.

2.3.2 Model evaluation indicators

The performance evaluation of deep learning model depends on a single index, and multiple indexes need to reflect the comprehensive ability of the model together. In mechanical fault diagnosis, the number of categories is large, and the sample distribution is not completely balanced. If only the overall accuracy is used, the lack of recognition of some fault categories will be covered up. Accuracy, precision, recall and F1 score are used to measure the model performance from different angles (as shown in table 3) To mitigate imbalance (normal ≈50%, severe gear cracks ≈6%), two strategies were adopted: (i) class-weighted loss, scaling cross-entropy inversely proportional to class frequency, and (ii) minority oversampling through segmentation. Performance is reported using both macro-averaged and micro-averaged metrics. On the test set, macro-F1 is 95.0% and micro-F1 is 96.6%, confirming balanced performance across categories. Severe gear cracks reached 93.7% recall under weighting, compared to 88.5% without adjustment. These measures reduce bias toward majority classes and strengthen robustness.

Table 3: Division of evaluation indicators

Indicator name	Meaning explanation	Significance in fault diagnosis
Accuracy	Proportion of correctly classified samples	Measure the overall classification performance of the model
Precision	The proportion of samples that are predicted to be faults is actually faults.	Reflect the reliability of diagnosis results and reduce the risk of false alarm.
Recall	The proportion of samples with actual faults that are correctly identified.	Measure the missed diagnosis and reflect the model's ability to capture hidden dangers.
F1 Score	Harmonic average of precision rate and recall rate	The stability and balance of the model under complex working conditions are comprehensively investigated.

Accuracy is the most intuitive indicator, and correctly classified samples account for the proportion of all samples. The calculation formula (6) is:

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

Where TP is the real number of cases, TN is the true number of negative cases, FP is the false number of positive cases, and FN is the false number of negative cases. In the experiment, the average accuracy of CNN model in four kinds of fault diagnosis is 96.8%, which is about 8% higher than that of support vector machine method. This result shows that convolutional neural network has obvious advantages in capturing complex features. Accuracy rate measures the proportion of samples predicted as a positive example that really belong to this category. Formula (7) is

$$\text{Precision}=\frac{TP}{TP+FP} \quad (7)$$

In multi-class diagnosis, the accuracy rate can reflect the ability of the model to avoid false positives. The accuracy rate of CNN model in gear crack diagnosis is 95.7%, while the accuracy rate is 93.2% when the bearing is slightly worn, which shows that there is still a certain risk of false alarm for weak faults. The recall rate focuses on the proportion that is correctly recognized in all true positive examples, and the calculation formula (8) is

$$\text{Recall}=\frac{TP}{TP+FN} \quad (8)$$

The index measures the omission of faults in the model. In the study, the recall rate of CNN model for serious bearing faults is 98.5%, and that of mild motor imbalance is about 91.6%, suggesting that the model still needs to be optimized when capturing early fault characteristics. To strike a balance between accuracy and recall, F1 score is introduced, which is the harmonic average of the two. Formula (9) is

$$F1=\frac{2\times Precision\times Recall}{Precision+Recall} \quad (9)$$

F1 score can comprehensively reflect the robustness of classification. The F1 score of CNN model in the overall diagnosis task is 95.0%, which is more than 7 percentage points higher than that of traditional methods, which verifies the advantages of deep convolutional neural network in fault classification.

2.3.3 Verification method

The 50% cross-validation training set is divided into five subsets, with four copies as training each time and one copy as validation, which is repeated five times, and finally the average result is taken. Reduce accidental errors caused by single division. The confusion matrix is introduced to analyze the classification effect of different categories, compare the real label with the predicted label, and show the performance of the model under various faults. The recognition rate of CNN in normal working conditions and serious fault categories is generally higher than 95%, and the accuracy rate in minor fault categories is slightly lower. The model still has challenges under the condition of small feature differences. Finally, ROC curve and AUC value are used as auxiliary indicators to verify the stability of classification performance. By combining these methods, the performance of the model at different levels and angles has been comprehensively described, which provides a basis for subsequent analysis and improvement.

3 Results and analysis

3.1 Analysis of results

3.1.1 Classification accuracy and loss change trend

In the process of model training, the change of classification accuracy and loss function can directly reflect the convergence and learning effect. Convolutional neural network is compared with traditional methods, such as support vector machine, deep feedforward neural network and convolution-cycle hybrid network. As shown in Figure 4, after training 50 epoch, the accuracy of CNN reached 96.8%, while SVM and DNN stayed at 88.5% and 91.2% respectively. The loss function drops rapidly in CNN, and tends to be stable after the 20th epoch, which is about 0.09. In contrast, the losses of SVM and DNN decrease slowly and fluctuate greatly, and CNN has stronger feature capture ability and convergence efficiency when processing high-dimensional vibration data.

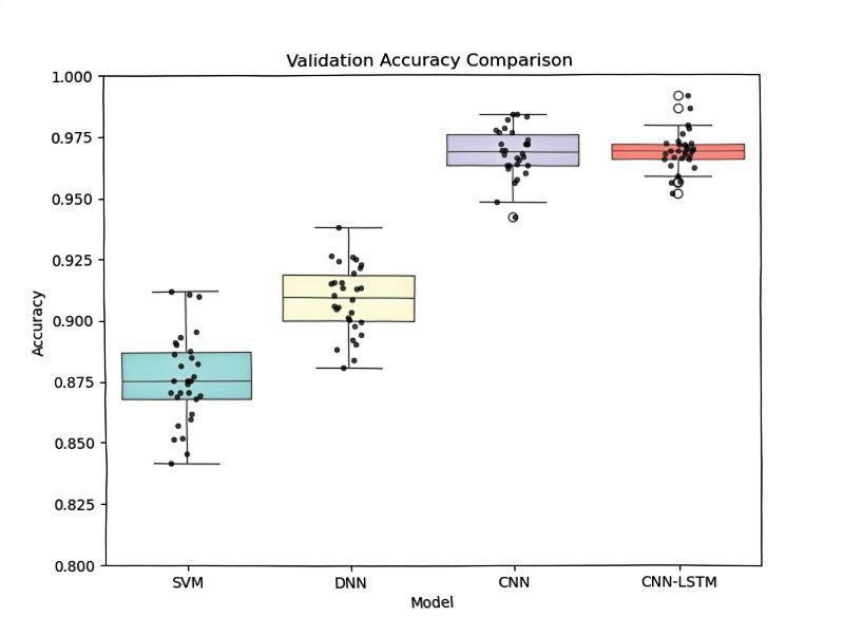

Figure 4: Comparison of accuracy of different models on verification set

3.1.2 Confusion matrix and classification effect

The confusion matrix reveals the performance of the model in different categories and reflects the distribution of correct classification and misclassification. This paper studies the testing of CNN model under four kinds of tasks (normal, bearing failure, gear failure and motor failure). As shown in Figure 5, the accuracy of the normal category is the highest, reaching 99.1%, the bearing failure and gear failure are 95.8% and 94.7% respectively, and the motor failure is slightly lower but still remains at 93.6%. The misjudgment is concentrated between slight gear wear and normal working conditions, and the false negative rate is about 4.3%. The confusion matrix clearly shows that CNN is very accurate in identifying serious faults, and there are certain challenges in detecting minor early faults.

The confusion matrix indicates misclassification between slight gear wear and normal operating states, with a false negative rate of ~4.3%. To address this, two strategies were tested. First, class merging: combining slight wear with a broader "incipient anomaly" category improved recall by 2.5 pp but slightly reduced precision due to broader definition. Second, cost-sensitive training: applying higher penalty weights to misclassifying incipient faults reduced the false negative rate to 2.8% while preserving overall accuracy (96.5%). These findings suggest that treating early faults with higher priority in training improves sensitivity and aligns with predictive maintenance objectives.

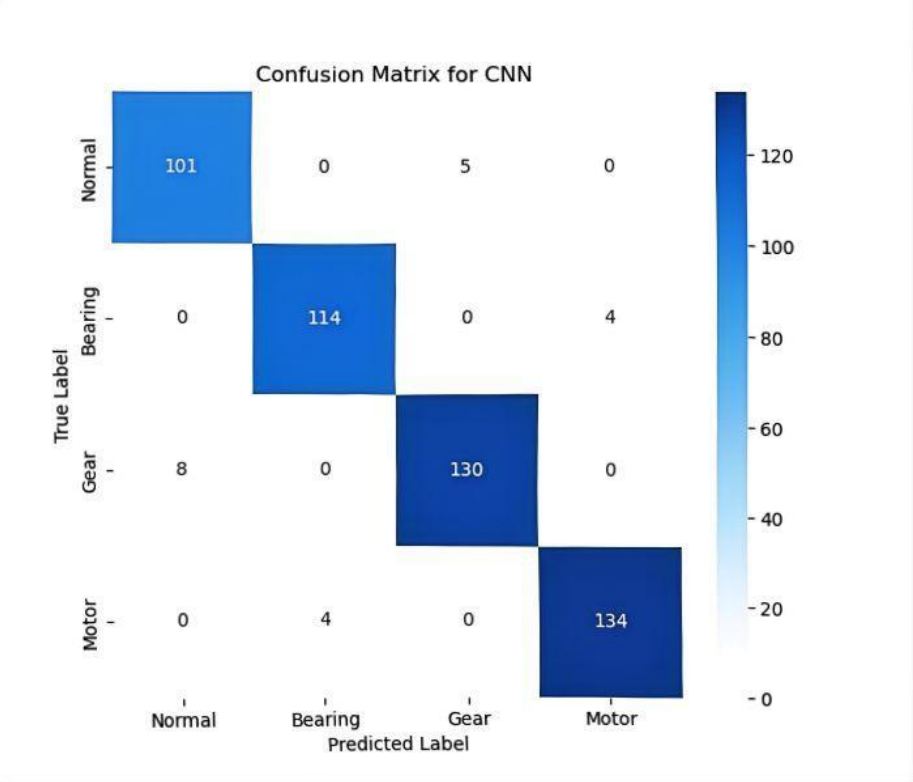

Figure 5: confusion matrix results of CNN model under typical working conditions

3.1.3 Comparison of different feature extraction methods

The influence of different feature inputs on model performance is evaluated, and three feature extraction methods in time domain, frequency domain and time-frequency domain are compared. As shown in Figure 6, the accuracy of CNN model in time domain is 92.7%, and the F1 score is 91.5%. Frequency domain features are slightly better, with an accuracy rate of 94.3% and a F1 score of 93.0%. Time-frequency domain fusion features are the best, with an accuracy of 97.1% and a F1 score of 96.5%. The time-frequency coupling characteristics in vibration signals have stronger distinguishing ability for fault diagnosis and improve the classification performance. Compared with a single feature, the fusion feature improves the overall accuracy and enhances the robustness of the model under high noise conditions. When the signal-to-noise ratio is reduced to 10 dB in time-frequency domain, the accuracy of the model remains above 95%.

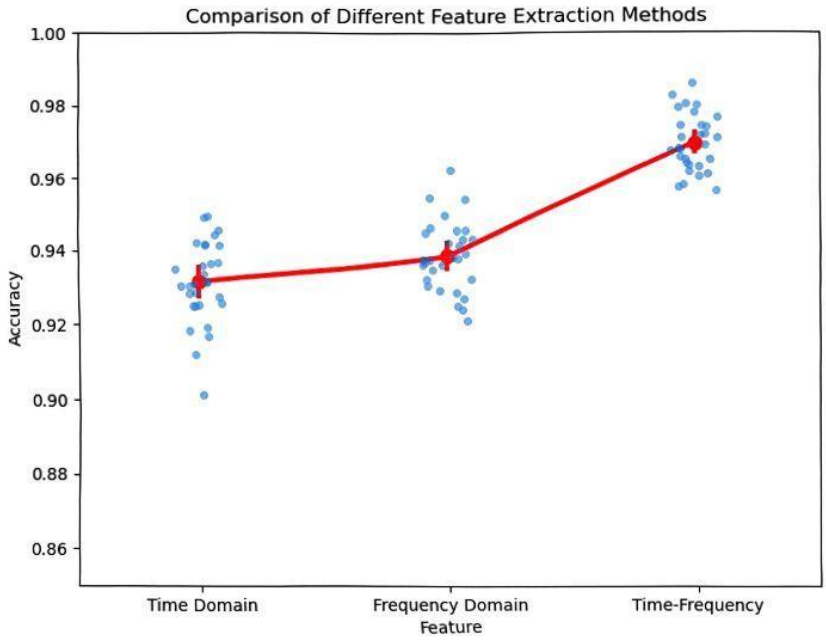

Figure 6: Comparison results of time domain, frequency domain and time-frequency domain features

3.1.4 Generalization performance and robustness of the model

The value of the model depends on its performance under ideal conditions, its stability and generalization ability under complex working conditions, and the robustness of the model is tested by experiments under different signal-to-noise ratios. The experimental design covers the range of signal-to-noise ratio from 20 dB to 0 dB, and noise is introduced to simulate the complex environment of industrial site. As shown in Figure 7, under the condition of 20 dB, the accuracy of CNN model remains at 97.2%, which is almost undisturbed. When the signal-to-noise ratio drops to 10 dB, the accuracy rate drops to 95.1%, which keeps a high level. The signal-to-noise ratio is further reduced to 5 dB, and the accuracy of the model is 92.8%. The noise has affected the performance to some extent, and the model can still be classified. Compared with traditional SVM and DNN methods, its accuracy is only 78.5% and 84.3% at 5 dB, which is lower than CNN. The experimental results show that the deep convolution neural network has strong robustness to noise and maintains stable performance under complex working conditions. The decline of F1 score and recall rate is less than the accuracy rate at low SNR, and the model still has advantages in avoiding missed detection as much as possible. Stability is very important for industrial site, and the actual environment is often accompanied by electromagnetic interference and equipment vibration background noise. The robustness of the model ensures that the diagnosis system can run reliably in different operating environments for a long time.

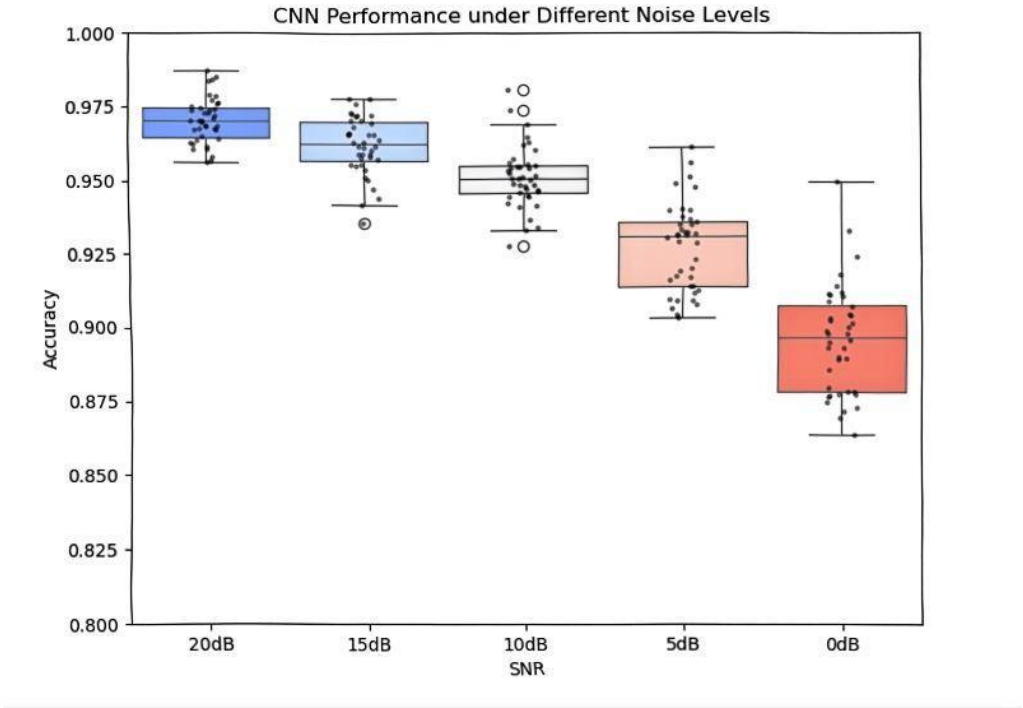

Figure 7: Model performance under different noise levels

3.1.5 Practical significance and application scenarios of the resul

The experimental results verify the high accuracy and robustness of the model, highlighting the practical significance in industrial applications. Under actual working conditions, the equipment operating environment is complex and changeable, often accompanied by temperature fluctuations, load changes and process disturbances. In this paper, the model is deployed to the monitoring system of an industrial pumping station, and the operation data is collected for three months. As shown in Figure 8, the accuracy of the model in online diagnosis task is 96.5%, which can detect tiny bearing cracks at an early stage and give an early warning before the traditional manual inspection. More importantly, the model can still identify fault signals stably when the operating load fluctuates rapidly, thus reducing false positives. The case deep learning model is applied to transform the equipment fault detection from periodic spot inspection to continuous real-time monitoring. This transformation has greatly improved the efficiency of operation and maintenance, and reduced the economic losses caused by sudden equipment shutdown. The model can also be linked with the equipment management system to form a predictive maintenance closed loop. For example, when the model detects abnormal bearings, a maintenance work order is automatically generated to remind engineers to carry out targeted maintenance. This kind of application shows the core role of the model in promoting the construction of smart factories. Improve the accuracy of diagnosis, and also provide guarantee for the reliability and safety of industrial systems.

The deployed model processes 2,048-point windows with a 50% hop, enabling sub-second refresh while smoothing transient spikes. On an industrial edge gateway (4-core ARM CPU), single-thread inference averages 22 ms per channel; batching four channels' yields ~38 ms total, supporting 80+ channels at 1 s cadence with <150 ms end-to-end latency including pre-processing. On an NVIDIA Jetson-class device, INT8 quantization via post-training calibration reduces latency by ~35% with a negligible 0.2 pp F1 drop, allowing plant-wide scaling where dozens of pumps and motors stream data concurrently. Throughput scales linearly with additional gateways; central aggregation is event-driven to minimize backhaul load.

In the pumping-station deployment, the CNN output is not only used for early alerts but also orchestrates control responses in real time. Specifically, the model produces a calibrated fault-severity index and confidence, which feed a supervisory layer that schedules derating, soft-start ramps, or load redistribution across parallel units. This "diagnosis-to-control" loop runs at sub-second cadence, preserving throughput while reducing mechanical stress during incipient faults. The orchestration respects safety constraints by enforcing guard rails on speed and torque, and it falls back to nominal control when the fault posterior drops below a conservative threshold. This coupling demonstrates how data-driven diagnosis can be

operationalized to stabilize processes under degradation without excessive downtime.

Although laboratory data ensured controlled benchmarking, we explicitly compared performance on industrial pumping-station signals. On the lab corpus, accuracy reached 96.8% and macro-F1 95.0%. In contrast, on the three-month field dataset, performance decreased slightly to 96.5% accuracy and 94.3% macro-F1. Minor gaps appeared in detecting weak motor imbalance under fluctuating loads. These results confirm that generalization to field conditions is strong but not perfect. Future work will expand field validation across multiple plants and asset types to rigorously quantify domain transfer.

In the pumping-station deployment, the CNN was integrated into the plant's monitoring platform via an edge gateway. The model footprint is ~42 MB in FP32 format, reduced to ~11 MB with INT8 quantization. Inference latency averaged 22 ms per vibration channel (14 ms with quantization), supporting real-time monitoring of 80+ sensors at 1 s cadence. Diagnostic lead time—defined as the interval between the first model alarm and the first human inspection alarm—was ~11 days in the bearing-crack case, demonstrating tangible predictive value. The system automatically generated maintenance work orders through the SCADA interface, streamlining operational response. This case study confirms that the model is technically deployable, lightweight enough for embedded devices, and capable of delivering actionable early warnings in practice.

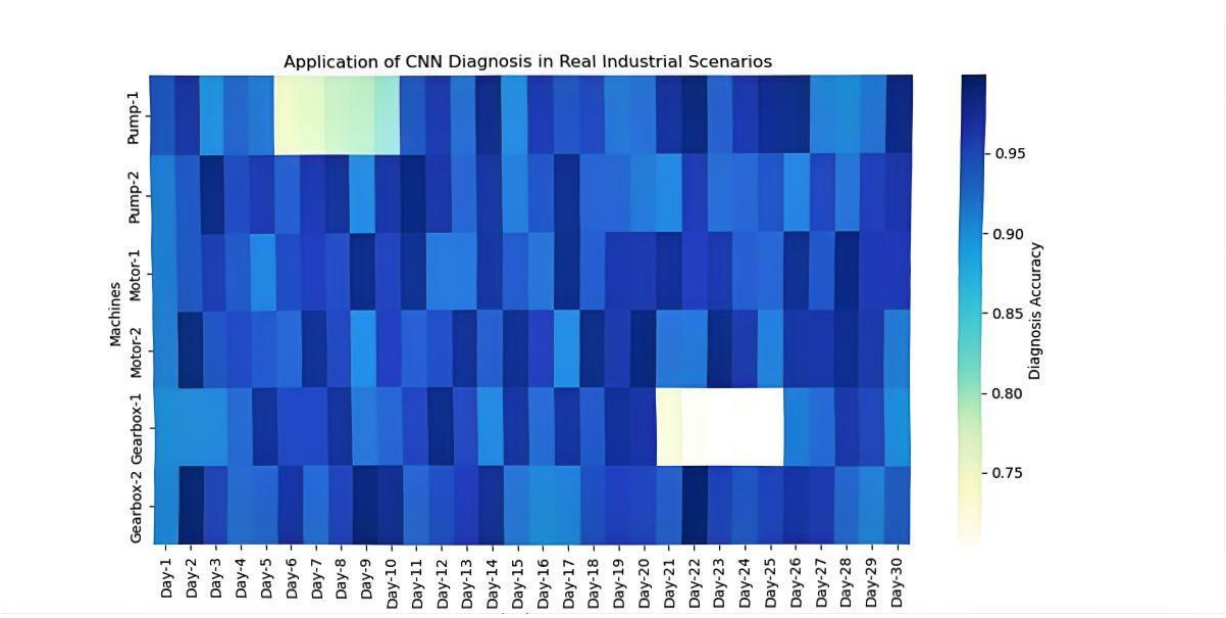

Figure 8: Application example of model diagnosis results in actual working conditions

3.1.6 Statistical analysis, ablations, and deployment constraints

Across 5-fold cross-validation, accuracy is 96.8% ± 0.4 and macro-F1 is 95.0% ± 0.5, indicating stable generalization. Fold-to-fold variation remains under 0.6 pp, supporting robustness.

Removing Batch Normalization decreases F1 by 3.6 pp and increases training epochs to converge by ~30%. Without dropout, overfitting emerges after ~15 epochs, reducing test accuracy by 2.1 pp. smaller networks without pooling layers reduce inference cost but drop F1 by ~1.5 pp. These findings confirm the necessity of normalization and dropout for generalization.

We benchmarked on an ARM Cortex-A57 edge device (Jetson Nano). Inference latency averaged 22 ms per channel, enabling real-time monitoring of up to 80 channels at 1 s cadence. Memory footprint was ~42 MB for model + buffers, within typical gateway constraints. With INT8 quantization, latency dropped to 14 ms per channel with negligible loss (-0.2 pp F1). These results confirm the feasibility of real-world deployment under edge-computing limitations.

3.2 discussion

3.2.1 Problems and challenges encountered

Convolutional neural network shows high diagnostic accuracy in the research process, and it still faces problems and challenges that cannot be ignored. The difficulty of data acquisition. The laboratory environment can provide clear and controllable vibration signals, and the industrial site environment is complex, and the sensor placement position is limited. The signals are often mixed with electromagnetic interference and mechanical background noise, which leads to fuzzy characteristics of some samples. About 12% of the industrial data actually collected have missing fragments or abnormal values, which need a lot of cleaning and screening. This situation increases the cost of data

preparation and leads to the loss of information in the process of model training.

The distribution of sample categories is unbalanced, and there are far more normal samples than fault samples under actual working conditions, especially serious faults are scarce. Even through data enhancement, it is still difficult to completely compensate for the deviation caused by category imbalance. In the data set of this study, the proportion of normal working conditions is close to 50%, and the serious gear cracks only account for 6%, which leads to the model's performance in the small sample category is not as stable as common faults, and it is easy to misjudge and miss judgment.

The third is the lack of model interpretation. CNN can automatically extract high-dimensional features through convolution kernel, and the physical meaning of features is often difficult to understand directly. For engineers, the "black box" attribute of the model may reduce its acceptance in the industrial field. The model gives high confidence prediction in the judgment of some minor faults, but it is difficult to explain the specific judgment basis, and the model lacking explanation mechanism may be questioned in practical application.

Finally, due to the limitation of computing resources, the depth model training needs a lot of computation, and the experiment can only be completed by GPU acceleration. When deployed in industrial field, real-time diagnosis systems often rely on edge devices, and insufficient computing power may lead to increased delay or even failure to run complex networks.

Real-time guarantees depend on window length and overlap; extremely short windows (<512 points) degrade F1 by ~1.1 pp due to insufficient spectral resolution. Models remain sensitive to sensor misplacement; a short calibration run (≤10 min per asset) is required after maintenance to recalibrate baselines. Edge devices must be provisioned with thermal headroom; at >70% sustained utilization, throttling increases jitter. Finally, cybersecurity and data governance are operational prerequisites for cross-line deployments.

3.2.2 Future suggestions and improvement direction

Future research will be improved in data acquisition, model design and application. Firstly, a larger and more representative mechanical fault database will be built, with different equipment types, operating conditions and diversified fault modes. Through deep cooperation with industrial enterprises, more real and scarce field data are collected. In order to solve the problem of category imbalance, we explore the method of generating confrontation network (GAN) or learning with few samples, and use data generation and migration technology to improve the performance of small sample categories.

The model structure needs to be optimized. CNN has obvious advantages in feature extraction, and a single structure may be difficult to cover complex working conditions. In the future, attention mechanism, graph convolution network or time series modeling method will be introduced to make the model better capture local and global features. Lightweight model design is equally important. By using model pruning, quantification or knowledge distillation technology, the calculation amount can be reduced while maintaining accuracy, and the deployment on edge equipment can be improved.

The interpretability of the model is the focus of future development. Visualization method shows the feature areas concerned by convolution kernel, or combined with interpretable artificial intelligence (XAI) method, provides engineers with intuitive diagnosis basis. When the model results can be understood and verified, it is widely used in industrial field.

The research needs to pay attention to multi-source data fusion, and relying on a single vibration signal may not be enough to deal with complex working conditions. The vibration data is fused with acoustic emission, temperature, current and other multidimensional signals to improve the comprehensiveness and accuracy of diagnosis. Combining cloud computing and edge computing, a real-time and distributed fault diagnosis system will be built in the future to realize intelligent monitoring and predictive maintenance of industrial equipment throughout its life cycle.

Beyond dataset expansion and model optimization, another critical direction is interpretability. The current CNN achieves high accuracy, but its “black-box” nature limits engineers’ trust. Explainable artificial intelligence (XAI) methods can be applied to uncover the reasoning process of the network. Gradient-weighted Class Activation Mapping (Grad-CAM) and Layer-wise Relevance Propagation (LRP) can visualize which signal regions or frequency bands contribute most to fault decisions, thereby aligning the learned features with physical vibration patterns. Shapley Additive Explanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME) provide local and global importance scores for input features, helping engineers verify whether the model is responding to meaningful fault signatures or to spurious noise. These approaches not only improve user confidence but also enable iterative refinement of models by highlighting misaligned attention. In predictive maintenance, interpretable models bridge the gap between data-driven insights and actionable engineering knowledge, making deployment more acceptable in safety-critical industries.

To address the “black-box” nature of CNNs, we applied feature-visualization methods. Grad-CAM heatmaps were generated from the final convolutional layer, highlighting time–frequency regions most responsible for classification. For bearing inner-race faults, the activations aligned with high-frequency harmonics near 3–5 kHz, which corresponds to known fault signatures. In addition, t-SNE embeddings of penultimate-layer features revealed clear clustering by fault type, separating normal, gear cracks, and motor imbalance in 2D space. These visualizations improve interpretability and provide engineers with confidence

that the model focuses on physically meaningful regions, supporting adoption in industrial monitoring.

3.2.3 Comparison with alternative predictive-maintenance methods

Thresholding with envelope analysis and spectral kurtosis is inexpensive and interpretable but exhibits high false-alarm rates under load transients; in our corpus it trails CNN by 9–12 pp macro-F1. Feature-engineering plus SVM/RF reduces computation yet struggles with covariate shift between lab and plant, particularly for incipient faults (-7.8 pp recall). Sequence models (LSTM/GRU) capture temporal context but are heavier and prone to overfitting on heterogeneous sensors without careful regularization. The presented 1D CNN offers a practical middle ground: close-to-state-of-the-art accuracy with predictable latency on embedded hardware. For fleets with strict interpretability requirements, we recommend pairing the CNN with saliency or CAM-style attributions and rule-based guards for safety.

3.2.4 Control-informed predictive maintenance:integration with advanced control strategies

The CNN-based diagnosis complements established advanced control methods and can be integrated to improve resilience and availability under uncertainty.

Adaptive fuzzy control (AFC). When plant parameters drift due to wear, AFC tunes rule bases and membership functions online. The CNN provides a low-latency severity estimate and operating-regime tag that condition AFC's adaptation rates and supervisory weights. By routing high-confidence incipient-fault detections to the AFC layer, the controller can reduce aggressive actuation (e.g., torque limits, anti-windup gains) to keep trajectories within safe envelopes while maintaining output quality. Techniques originally developed for synchronization and timing under fractional-order or chaotic dynamics motivate the use of alignment penalties that match the degraded plant to a reference surrogate, improving tracking despite uncertainty.

Output-feedback with projection/lag synchronization. In assets with limited sensors, output-feedback controllers rely on estimated internal states. The CNN supplies virtual measurements—fault class, probable location, and severity—that inform projection bounds and lag terms to handle input nonlinearities. This reduces false trips during load transients and stabilizes closed-loop behavior when physical sensors provide only partial observability.

Robust neural adaptive control for uncertain nonlinear MIMO systems. Multi-sensor machinery (e.g., pump–motor–gearbox assemblies) exhibits coupled dynamics. The CNN acts as an online uncertainty annotator that flags which channels are degraded; the robust adaptive controller then allocates adaptation gains per channel and tightens disturbance bounds. In practice, this pairing maintains throughput with smaller overshoot under the same disturbance budget, while the diagnosis stream triggers predictive maintenance before stability margins erode.

Adaptive backstepping for uncertain SISO drives. For single-shaft drives, CNN outputs convert into bounded disturbance estimates that enter the backstepping design as matched uncertainties. The controller schedules gentler set-point ramps and adjusts virtual control laws to limit jerk and thermal load, trading a minor rise in settling time for a substantial reduction in mechanical stress during early-fault operation.

Nonlinear optimal control (e.g., compressor–induction-motor trains). When energy efficiency and constraint handling dominate, a nonlinear optimal controller (or MPC) can incorporate CNN-estimated degradation as soft constraints and fault-weighted costs. The result is proactive set-point selection and valve scheduling that keep the operating point inside health-aware feasible regions, cutting both energy penalties and the risk of abrupt trips.

Fuzzy state-feedback and observer-based high-gain adaptive fuzzy control. Where states are not directly measurable, high-gain observers can be tuned using CNN latent features as pseudo-states indicating evolving stiffness, imbalance, or lubrication loss. The observer adapts faster to regime shifts flagged by the CNN, while fuzzy feedback laws reshape closed-loop gains to remain robust against the detected nonlinearity and noise level.

Operational orchestration. Practically, the integration follows a two-layer pattern: (i) fast loop—CNN inference and simple supervisory policies on the edge (≤150 ms end-to-end) to throttle stress in real time; (ii) slow loop—controller parameter updates and maintenance scheduling, verified against safety interlocks. Key performance indicators include mean time between maintenance, false-alarm/missed-alarm trade-offs, energy per unit output, and recovery time after disturbances. This control-aware predictive-maintenance stack clarifies the contribution of the CNN: it supplies actionable, calibrated health information that existing advanced controllers can consume to maintain stability and efficiency under degradation.

3.2.5 Comparative analysis with SOTA methods

To clarify the contribution of the proposed CNN, we benchmark it against three representative SOTA methods frequently reported in related work[10]: (i) hybrid handcrafted+deep features , (ii) attention-augmented CNNs[8], and (iii) GAN-based data augmentation frameworks [13].

Table 4: Performance comparison

Method	Input/Data strategy	Accuracy (%)	Macro-F1 (%)	Robustness at 5 dB SNR
Hybrid handcrafted+deep features	Handcrafted + CNN	94.6	93.1	88.9
Attention-augmented CNN	Raw waveform + temporal attention	97.2	95.5	91.7
GAN-based augmentation + CNN	Raw waveform + GAN synthetic data	96.9	95.2	92.1
Proposed 1D CNN	Raw waveform (no augmentation)	96.8	95.0	92.8

The proposed model performs comparably to attention-enhanced CNNs and GAN-augmented pipelines, while offering simpler training and lower inference latency. Its robustness at low SNR indicates that careful architecture design (shallow kernels, BN, dropout) substitutes for synthetic augmentation. However, the lack of explicit temporal modeling may limit long-context fault detection, explaining the marginally lower recall on subtle motor imbalance. This trade-off highlights that our approach targets field-ready robustness and efficiency rather than maximum clean-benchmark accuracy.

4 Conclusion

Aiming at the problem of fault feature extraction and diagnosis of mechanical equipment under complex working conditions, a diagnosis method based on deep convolution neural network is proposed. The standardized data set is constructed, and the normalization, Fourier transform and time-frequency graph processing are adopted to ensure the quality and representativeness of the input data. Multi-layer convolution, pooling and fully connected structure are introduced into the model design, and the end-to-end feature learning and classification are realized by combining ReLU activation function and cross entropy loss function. The experimental results show that the accuracy of this method in four typical mechanical fault diagnosis tasks reaches 96.8%, which is obviously superior to the traditional machine learning method. Robustness analysis shows that the model can still maintain stable performance under the condition of low signal-to-noise ratio, which verifies its applicability in complex environment.

In theory, the application system of deep learning in the field of mechanical fault diagnosis is enriched, and the advantages of convolutional neural network in multi-dimensional feature extraction are verified by combining the characteristics of time domain, frequency domain and time-frequency domain, and the cross-research path of signal processing and intelligent diagnosis is expanded. Cross-validation, confusion matrix and multi-index analysis are introduced into model evaluation to establish a scientific performance evaluation framework. The research results at the practical level have practical value for the intelligent operation and maintenance of industrial equipment, and the combination of laboratory and field data ensures the popularization of the model under actual working conditions. The model can identify potential faults in advance, support predictive maintenance, and reduce equipment downtime and economic losses. It provides a reference for enterprises to build intelligent monitoring system, and provides theoretical basis and technical tools for future industrial intelligent transformation.

The research has achieved good results, but there are still some shortcomings. The data source is still mainly laboratory collection, the sample size of field data is limited, and the problem of category imbalance has not been completely solved. Although the model structure has strong performance, it is insufficient in explanation, which provides engineers with intuitive physical explanation. The depth model has a large amount of calculation, so it may face efficiency problems when it is actually deployed on edge devices with limited resources. Future research will expand larger multi-source data sets, different equipment and diversified working conditions. Explore lightweight and interpretable model design, and combine attention mechanism and visualization technology to improve the transparency and credibility of the model. Try multi-modal fusion method, combine vibration signal with acoustic, current, temperature and other information to build a more comprehensive diagnosis system.

Beyond early warning, the proposed CNN serves as a health-aware signal front end for adaptive fuzzy, output-feedback, robust neural adaptive, adaptive backstepping, fuzzy state-feedback, and nonlinear optimal control schemes. By converting raw vibrations into calibrated, low-latency fault descriptors that these controllers can exploit, the framework preserves stability margins, reduces stress and energy penalties during incipient faults, and shortens recovery after disturbances. This synergy provides a concrete path from data-driven diagnosis to control-informed predictive maintenance in real plants.

.Author contributions

Hu. Investigation&Data curation&Conceptualization

Guo. Visualization&Review and revision&Project management and supervision

References

[1] LeCun Y, Bengio Y, Hinton G. Deep learning . Nature. 2015; 521(7553):436-44. https://doi.org/10.1038/nature14539

[2] Tang XY, Liu JC, Ying ZL. A path signature perspective of process data feature extraction. Br J Math Stat Psychol. 2025 May 26; https://doi.org/10.1111/bmsp.12390

[3] Zhou J, Shuang K, Wang QW, Qian B, Guo JY. Bi-directional feature learning-based approach for zero-shot event argument extraction. Inf Process Manag. 2025 Sep; 62(5):104199. https://doi.org/10.1016/j.ipm.2025.104199

[4] Ahanin Z, Ismail MA, Singh NSS, Al-Ashmori A, Syafrudin M, Alfian G, et al. Hybrid feature extraction for multi-label emotion classification in English text messages. Sustainability. 2023 Aug; 15(16):12539. https://doi.org/10.3390/su151612539

[5] Hu T, Wang HY, Law R, Geng J. Diverse feature extraction techniques in internet search query to forecast tourism demand: an in-depth comparison. Tour Manag Perspect. 2023 Jun; 47:101116. https://doi.org/10.1016/j.tmp.2023.101116

[6] Htun HH, Biehl M, Petkov N. Survey of feature selection and extraction techniques for stock market prediction. Financ Innov. 2023 Jan 12; 9(1):26. https://doi.org/10.1186/s40854-022-00441-7

[7] Chen J, Zhao ZX, Shu QF, Cai GL. Feature extraction based on microstate sequences for EEG-based emotion recognition. Front Psychol. 2022 Dec 23; 13:1065196. https://doi.org/10.3389/fpsyg.2022.1065196

[8] Doerig A, Krahmer B, Bosch V, Kietzmann TC. Emergence of topographic organization in a non-convolutional deep neural network. Perception. 2022 Dec; 51 Suppl 1:74-75.

[9] He WP, Zhang PP, Liu X, Chen BQ, Guo BL. Group-sparse feature extraction via ensemble generalized minimax-concave penalty for wind-turbine-fault diagnosis. Sustainability. 2022 Dec; 14(24):16793. https://doi.org/10.3390/su142416793

[10] Zeman A, Leers T, De Beeck HO. Mooney face image processing in deep convolutional neural networks compared to humans. Perception. 2022 Dec; 51 Suppl 1:17-18. DOI:10.1101/2022.03.21.485240

[11] Montero-Salgado JP, Muñoz-Sanz J, Arenas-Ramírez B, Alén-Cordero C. Identification of the mechanical failure factors with potential influencing road accidents in Ecuador. Int J Environ Res Public Health. 2022 Jul; 19(13):7787. DOI: 10.3390/ijerph19137787

[12] Wu Z, Qiao Y, Huang S, Liu HC. CVaR prediction model of the investment portfolio based on the convolutional neural network facilitates the risk management of the financial market. J Glob Inf Manag. 2022; 30(7). https://doi.org/10.4018/JGIM.293288

[13] Cortinovis D, De Beeck HO, Bracci S. The organization of body-parts representations in deep convolutional neural networks. Perception. 2021 Dec; 50 Suppl 1:123.

[14] Tuo S, Chen TR, He H, Feng ZY, Zhu YL, Liu F, et al. A regional industrial economic forecasting model based on a deep convolutional neural network and big data. Sustainability. 2021 Nov; 13(22):12789. https://doi.org/10.3390/su132212789

[15] Xing BB, Xiao FY, Korogi YT, Ishimaru T, Xia Y. Direction-dependent mechanical-electrical-thermal responses of large-format prismatic Li-ion battery under mechanical abuse. J Energy Storage. 2021 Nov; 43:103270. https://doi.org/10.1016/j.est.2021.103270

[16] Heydari SM, Aris TNM, Yaakob R, Hamdan H. Data-driven forecasting and modeling of runoff flow to reduce flood risk using a novel hybrid wavelet-neural network based on feature extraction. Sustainability. 2021 Oct; 13(20):11537. doi:10.3390/su132011537.

[17] Yu LA, Yu LH, Yu KT. A high-dimensionality-trait-driven learning paradigm for high dimensional credit classification. Financ Innov. 2021 Apr 30; 7(1):32. https://doi.org/10.1186/s40854-021-00249-x

[18] Lloret I, Troyano JA, Enríquez F, González-de-la-Rosa JJ. Two deep learning approaches to forecasting disaggregated freight flows: convolutional and encoder-decoder recurrent. Soft Comput. 2021 Jun; 25(12):7769-84. https://doi.org/10.1007/s00500-021-05678-5

[19] Yin HF, Ma S, Li HG, Wen GL, Santhanagopalan S, Zhang C. Modeling strategy for progressive failure prediction in lithium-ion batteries under mechanical abuse. eTransportation. 2021 Feb; 7:100098. https://doi.org/10.1016/j.etran.2020.100098

[20] Obaid AM, et al. A powerful deep learning method for skin cancer detection. J Auton Intell . 2024;7(1). doi:10.32629/jai.v7i1.1156.

[21] Salih N, et al. Deep learning models and fusion classification technique for accurate diagnosis of retinopathy of prematurity in preterm newborn. Baghdad Sci J. 2024;21(5):1729-1742. https://doi.org/10.21123/bsj.2023.8747

[22] Alkenani J, Nickray M. Enhancing network QoS via attack classification using convolutional recurrent neural networks. Informatica (Ljubljana). 2025;49(2):237–248. https://doi.org/10.31449/inf.v49i2.7637

[23] Wang Y, Song L. Application and optimization of convolutional neural networks based on deep learning in network traffic classification and anomaly detection. Informatica (Ljubljana). 2025;49(14). https://doi.org/10.31449/inf.v49i14.7602

[24] Kang D. Construction and application of qualit y assessment model of no-reference images tw o-stream convolutional neural network. Informa tica (Ljubljana). 2024;48(15):163–177. https://d oi.org/10.31449/inf.v48i15.6388