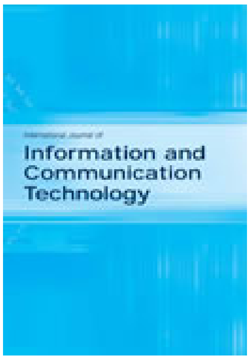


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Intelligent fault diagnosis system for railway infrastructure based on deep learning

Qi An

DOI: 10.1504/IJICT.2025.10074513

Article History:

Received:	31 August 2025
Last revised:	26 September 2025
Accepted:	28 September 2025
Published online:	20 November 2025

Intelligent fault diagnosis system for railway infrastructure based on deep learning

Qi An

Signal and Communication Research Institute,
China Academy of Railway Science Corporation Limited,
Beijing 100081, China
Email: anqi0573@163.com

Abstract: The operational status of railway infrastructure determines the safety of train passage. However, traditional research suffers from low efficiency and difficulty in addressing fault states under variable operating conditions. To address this, this paper first proposes a data balancing method based on improved synthetic minority over-sampling technique and generative adversarial network (GAN) to tackle the imbalance in railway infrastructure signal data. The introduction of unsupervised clustering algorithms and natural neighbour concepts enhances sample generation efficiency. Adding category label information and optimising the training loss function improves the stability of network training. Building upon this foundation, a multi-scale residual network (ResNet) is constructed for feature extraction, mitigating the impact of operational variations on diagnostic outcomes. A subdomain-adaptive transfer learning strategy is employed to achieve fault diagnosis. Experimental validation demonstrates that the proposed method achieves a diagnostic accuracy of 93.86%, delivering highly precise diagnostic results.

Keywords: railway infrastructure; fault diagnosis; synthetic minority over-sampling technique; generative adversarial network; GAN; transfer learning.

Reference to this paper should be made as follows: An, Q. (2025) 'Intelligent fault diagnosis system for railway infrastructure based on deep learning', *Int. J. Information and Communication Technology*, Vol. 26, No. 41, pp.107–122.

Biographical notes: Qi An received a Master's degree from China Academy of Railway Science in 2010. He is now a researcher at China Academy of Railway Science Corporation Limited. His main research direction is the high-speed railway earthquake early warning, intelligent diagnosis and intelligent fault handling.

1 Introduction

As the main arteries of national transportation, the safe and stable operation of railway infrastructure directly relates to economic development and social progress. Key facilities such as tracks, bridges, tunnels, catenary systems, and signalling equipment are highly prone to fatigue damage, deformation, aging, and sudden failures. Traditional fault diagnosis primarily relies on manual inspections and periodic maintenance, which is not only inefficient and costly but also significantly affected by subjective experience, making it difficult to achieve precise perception and early warning throughout all conditions and coverage areas (Hu et al., 2022). Deep learning can automatically learn complex features and patterns from large amounts of data, offering new ideas and approaches for solving railway infrastructure fault diagnosis problems (Han, 2024). Intelligent fault diagnosis systems based on deep learning can collect operational data in real-time, perform in-depth analysis and mining through deep learning algorithms, enabling automatic detection, classification, and localisation of faults. This greatly improves diagnostic accuracy and efficiency (Chen et al., 2021). Consequently, the investigation of intelligent fault diagnosis systems for railway infrastructure utilising deep learning holds both profound theoretical significance and substantial practical value (Wang et al., 2024).

Early research mainly conducted railway infrastructure fault diagnosis based on traditional signal analysis (Song et al., 2021). Ghosh et al. (2022) put forward a spectral kurtosis approach in light of wavelet packet decomposition and manifold learning, which realises the fault impact characteristics of enhanced signals by suppressing noise in time-frequency space. Sun et al. (2022) first decomposed bearing signals using wavelet transform, then used the spectral kurtosis method to find optimal frequency bands, and validated. Huang et al. (2019) proposed a variational mode decomposition (VMD) based on correlation analysis for adaptively extracting weak faults and composite fault characteristics of railway infrastructure, achieving good diagnostic results. Wu et al. (2023) proposed an adjustable factor-based wavelet transform method for high-speed train fault diagnosis, utilising quality factors to screen sub-band components, thereby enabling the separation of fault impact responses under strong interference. The operating environment of railway infrastructure is complex, involving various interfering factors such as electromagnetic interference, mechanical noise, and environmental vibrations. These interfering signals can mix with normal operational signals, affecting the diagnostic accuracy of traditional signal analysis methods.

Machine learning is a data-driven approach that realises data mining techniques, serving as an alternative strategy for relevant scholars after the failure of rule-based intelligent systems. It improves machine computing performance by extracting knowledge from massive datasets through machines. Hu et al. (2017) designed a railway infrastructure fault diagnosis approach based on the combination of genetic algorithms (GA) and backpropagation neural network (BPNN) to enhance its classification performance. Shao et al. (2020) developed an approach that employed multi-scale sample entropy to capture the features of vibration signals corresponding to various health states of railway assets. This method integrated a support vector machine (SVM) classifier, optimised via particle swarm optimisation, for the diagnosis and classification of infrastructure faults. Sun et al. (2023) extracted train bearing vibration signal features through morphological pattern spectrum and applied least squares SVMs for fault identification. The diagnosis accuracy was only 79.4%. Zhang et al. (2021) used principal component analysis (PCA) to reduce the dimensionality of temperature data features from train bearings, and proposed a method combining analytic hierarchy process with random forest algorithms for facility fault recognition. The aforementioned methods have gained good outcome in railway infrastructure fault diagnosis. However, machine learning algorithms still require manual involvement in the feature extraction section. They are overly reliant on signal analysis techniques and diagnostic experience. In particular, the labour-intensive feature extraction process makes the entire algorithm excessively complex, and the extracted fault features are often unsatisfactory.

The rise of deep learning has provided a solution to this, as it leverages its end-to-end learning capability to automatically extract features and build models from large amounts of collected data, and can adapt to different network structures such as convolutional, recurrent, or encoder-decoder mechanisms, demonstrating superior performance in device status identification and significantly simplifying the fault diagnosis process. Simone et al. (2023) suggested a convolutional long short-term memory network (LSTM) model for fault diagnosis by analysing signals from railway infrastructure; however, insufficient spatial feature extraction led to low diagnostic accuracy. Lv et al. (2024) combined deep residual network (ResNet) with transformer and proposed an algorithm for railway infrastructure fault diagnosis under variable working conditions, achieving a diagnostic accuracy of 81.9%. Shao et al. (2018) proposed a railway facility fault diagnosis method based on Hilbert-Huang transform deep feature representation and transfer learning to accurately predict the lifespan of facilities. Since railway infrastructure operates under normal working conditions, excessive diagnosis is not cost-effective. As initial fault signals are often weak and cannot be extracted due to noise, obtaining high-quality labelled fault data is very difficult. Generative models can capture statistical features of samples by fitting their underlying probability distribution, enabling the generation of pseudo-samples that are similar to and follow the same distribution as the data. Raza et al. (2025) expanded the fault dataset using conditional generative adversarial networks (GANs), achieving good results in multi-working condition diagnosis tasks, which verified the robustness of GAN under complex working conditions. Men et al. (2025) introduced deep convolutional GAN to enhance feature generation and effectively improved the model's recognition capability under extremely limited sample conditions, achieving a diagnostic accuracy as high as 91%.

In summary, traditional railway infrastructure fault diagnosis systems suffer from low efficiency, poor accuracy, and high costs, making it difficult to meet the standards of modern intelligent maintenance for railway equipment. Additionally, the scarcity of railway infrastructure fault data and imbalanced dataset categories often lead to low diagnostic accuracy in existing research. To address these issues, this article suggests an intelligent fault diagnosis system for railway infrastructure in light of deep learning. This system can fully address the issue of low diagnostic efficiency in existing research. The main contributions of this study can be summarised into four aspects.

1 To cope with the issue of data imbalance, a data balancing method is proposed based on an improved Synthetic Minority Over-sampling Technique (SMOTE) algorithm (ESMOTE) and an improved GAN (CSGAN). Initial railway infrastructure signal data are input into the ESMOTE method to generate one-dimensional vibration data across different states. Simultaneously, time-frequency image data of railway facility signals are fed into the CSGAN model to generate time-frequency images for various states, providing high-quality data support for this study.

2 ESMOTE achieves effective subdivision of intra-class data by introducing K-means clustering to make an initial division based on the inherent characteristics of each class. Simultaneously, it combines the concept of natural neighbours for linear interpolation to generate new samples, significantly reducing reliance on expert experience. CSGAN introduces category labels as additional information and uses

their corresponding embeddings as inputs to both the generator and discriminator, thereby constraining the data generation process to follow the form of target classes, generating samples that conform to specific class distributions.

3 A transfer learning approach is presented, which transfers the knowledge acquired from source domain data to the target domain, enabling effective identification of railway infrastructure faults under complex operating conditions. By constructing a feature extractor combining ResNet and multi-scale feature fusion technology, it enhances the network's ability to extract features, reduces the impact of working condition variations on diagnostic results, and also addresses the problem of network degradation caused by increasing depth. Furthermore, adopting a sub-domain adaptation transfer learning strategy enables railway infrastructure fault diagnosis.

4 A large number of simulation experiments were conducted on real datasets. The outcome implies that the suggested approach enhances fault diagnosis accuracy by 1.27%–3.4%, effectively enabling failure identification and classification for railway infrastructure under complex working conditions, providing a new approach to railway infrastructure fault diagnosis in scenarios with scarce sample data.

2 Relevant theory

2.1 Residual networks

ResNet is a revolutionary convolutional neural network framework in the field of deep learning. The central concept behind it lies in mitigating the issues of gradient vanishing and network degradation through the introduction of residual connections, allowing the network to be trained to unprecedented depths while maintaining excellent performance. Residual units are the key component of ResNet, enabling direct information transmission through added cross-layer pathways (Wu et al., 2019). It includes convolutional operations and skip connections, which allow input to bypass certain layers directly.

Assume that the learning target is $f(x)$. Among them, x is the shallow output. The stacked convolutional layers within the dashed box are newly added layers that learn residual mapping $f(x) - x$. Their summation produces the ideal mapping $f(x)$ at output. If it is desired to train the newly added level as an identity mapping $f(x) = x$, then simply set the weights and biases of the stacked convolutional layers in the residual block to zero. Therefore, the training error rate of the new network will be no worse than that of shallow networks, and using shortcut connections within the residual blocks can speed up data forward propagation and gradient backward propagation in neural networks, thus accelerating network training. In addition, this structural design requires the feature map shapes on the main path and the branch path in the residual block to be identical so that an addition operation can be performed at the output of the residual block (Fang et al., 2021).

2.2 Generative adversarial network

GAN includes a generator (G) and a discriminator (D), whose main idea is to continuously improve the performance of each network during training, so that G's fabricated data can fool the discriminator, while D strives to distinguish the authenticity of input data (Wang et al., 2025). In practice, G takes a random variable z conforming to a certain prior distribution $p_z(z)$ as input and generates pseudo-samples $G(z)$ based on it. The discriminator receives real samples and generated samples and outputs a probability value indicating that the sample is 'real'. Usually, if the output is greater than 0.5, it is judged as real; otherwise, it's fake. Through continuous training, the discriminator gradually improves its accuracy in identifying synthetic samples, while the generator makes its output data distribution $p_z(z)$ approach the true data distribution $p_{data}(x)$, finally reaching a Nash equilibrium where any data has a probability of being considered real as 1/2. The training and optimisation process of the discriminator and generator can be viewed as a maximisation-minimisation problem, with specific functions shown in equation (1).

$$\min_{G} \max_{D} V(D, G) = E_{x \sim p_{pata(x)}} \left[\log D(x) \right] + E_{z \sim p_{z(z)}} \left[\log \left(1 - D\left(G(z) \right) \right) \right] \tag{1}$$

where $V(D, G)$ is the final optimisation result of GAN, E represents mathematical expectation. When the structure of the discriminator is fixed, the training objective of the generator is to minimise this function to generate more deceptive samples. When the structure of the generator is fixed, the discriminator strives to maximise this target function to enhance its ability to distinguish between real and fake samples.

2.3 SMOTE oversampling algorithm

Traditional random oversampling methods simply duplicate minority class samples without adding new information. For example, in a dataset containing two classes, if the minority class has few samples, random oversampling may repeatedly select existing minority samples to add to the training set. This leads to a large number of duplicate samples in the training data. During learning, the model may become overly reliant on these duplicates, making it excessively sensitive to noise and specific patterns within the training data. Ultimately, this causes overfitting, resulting in poor generalisation ability when encountering new, unseen data. The SMOTE algorithm does not simply duplicate minority class samples. Instead, it synthesises new samples by performing linear interpolation of minority class samples in the feature space. This approach increases sample diversity, providing the model with more varied information. It helps the model learn more general features and patterns, thereby reducing the risk of overfitting and improving the model's generalisation ability. The SMOTE oversampling algorithm generates synthetic sample sets that conform to the original data distribution by performing oversampling interpolation operations on minority class samples in the feature space, further alleviating the problem of class imbalance (Duan et al., 2022). This mechanism effectively reduces training bias in classification models and significantly improves model generalisation performance. The process of generating new samples through SMOTE is based on the distances between a sample itself and its k nearest neighbours. Common distance metrics include Euclidean distance, Manhattan distance, etc. (Elreedy et al., 2024). The SMOTE algorithm operates according to the following procedure.

Firstly, take a certain sample x_i from the minority class sample set as a benchmark, calculate the distances between this sample and other samples using Euclidean distance, and select k nearest neighbours based on the calculation results. Secondly, randomly select one of these k nearest neighbour samples x_{ij} as an auxiliary sample, and perform linear interpolation with x_i according to equation (2) to obtain a new generated sample x_{new}, where α is a random number between [0, 1].

$$x_{new} = x_i + \alpha \times (x_{new} - x_i) \tag{2}$$

Finally, set the sampling rate N based on the sample imbalance ratio, repeat the aforementioned process N times to derive all new instances, and effectively enhance the class distribution of the dataset.

3 Railway infrastructure signal data balancing based on improved SMOTE oversampling and GAN

3.1 Design of the improved SMOTE algorithm

Due to the scarcity of railway infrastructure fault data and imbalanced dataset classes leading to poor model training performance and low diagnostic accuracy, this paper proposes an ESMOTE and CSGAN-based data balancing method based on data characteristics. ESMOTE combines unsupervised clustering algorithms with natural neighbour concepts, while CSGAN introduces class label information and optimises the training loss function to enhance the quality of produced samples across different datasets.

During the data balancing process, over-synthesis or noise introduction can affect the quality of generated samples. Although SMOTE is a widely used classical oversampling method, it has two main limitations. First, choosing the nearest neighbour parameter k heavily relies on expert experience and lacks adaptive adjustment capability. Second, it does not comprehensively consider the distribution characteristics of the data itself, making it likely to generate samples inconsistent with actual data distributions, hence influencing the performance of subsequent models. To tackle these problems, this paper presents an enhanced ESMOTE algorithm, which operates according to the following procedure.

First, introduce the K-means clustering algorithm (Ahmed et al., 2020) to preprocess minority class data sequentially and divide it into c sub-clusters with similar feature distributions. This relationship is represented by $D_s = \{D_{s1}, D_{s2}, D_{s3}, \ldots, D_{sc}\}$, where D_s denotes the minority class data, D_{sj} represents the j-th cluster, and $j \in [1, c]$.

Second, randomly select base samples $x_{Base}^{(i)}$ from each sub-cluster and randomly choose auxiliary samples $x_{Nearest}^{(i)}$ within their neighbouring sample sets in light of natural neighbour relations. Compared to the k-nearest neighbours method used by traditional SMOTE algorithms, natural neighbour relationships offer stronger adaptability (Zhu et al., 2016). The stable definition of natural neighbour relationships is as follows: if sample p is a natural neighbour of sample q, then sample p is also one of the λ nearest neighbours of sample q; similarly, sample q is also one of the λ nearest neighbours of sample p, where λ represents the characteristic value of natural neighbours.

$$p \in NaN(q) \Leftrightarrow q \in NN_{\lambda}(p) p \in NN_{\lambda}(q) \tag{3}$$

Finally, set a sampling rate N based on the imbalance ratio of the dataset and generate new samples between $x_{Base}^{(i)}$ and $x_{Nearest}^{(i)}$ using linear interpolation. The calculation expression is as follows: $x_{Base}^{(i)}$ represents the i^{th} base sample in the cluster; $x_{Nearest}^{(i)}$ represents the auxiliary sample of $x_{Base}^{(i)}$, diff denotes the difference between $x_{Nearest}^{(i)}$ and $x_{Base}^{(i)}$, $x_{New}^{(i)}$ is the synthesised new sample, and α is a random factor with $\alpha \in [0, 1]$.

$$diff = x_{Nearest}^{(i)} - x_{Base}^{(i)} \tag{4}$$

$$x_{New}^{(i)} = x_{Base}^{(i)} + \alpha \times diff \tag{5}$$

ESMOTE divides each class into effective subgroups by introducing K-means clustering according to intrinsic data characteristics. Meanwhile, it combines natural neighbour concepts for linear interpolation to generate new samples, significantly reducing dependence on expert experience. Compared to traditional SMOTE algorithms, ESMOTE ensures a reasonable intra-class distribution and controls the quality of generated samples during the sampling process, further demonstrating its advantages in data balancing tasks.

3.2 GAN model improvement and railway infrastructure signal data balancing process

During the training process of GANs, when cross-entropy is chosen as the loss function, if there are significant differences between the real data distribution and the generated distribution, the discriminator tends to make erroneous judgments, leading to an unstable model training process and poor quality of the generated samples. Conditional GAN uses least squares mean square error instead of traditional cross-entropy measures by constructing adversarial regularisation terms based on decision boundaries, forcing the generated samples in the feature space to align with the real data distribution judged by the discriminator, thus improving the quality of the generated samples (Abu-Srhan et al., 2022).

However, the generator of conditional GAN still has issues such as limited output sample diversity and insufficient condition control sensitivity. To further optimise model performance, this paper proposes a Conditional Least Squares GAN (CSGAN). First, it introduces class labels as additional information and incorporates their corresponding embeddings into the inputs for both the generator and discriminator to constrain the data generation process to follow target classes, thus generating samples that conform to specific class distributions. Second, an $L2$ regularisation term is added to the training loss function of the discriminator to impose reasonable penalties on parameters, enhancing parameter distribution uniformity, effectively improving the generalisation ability of the discriminator, and preventing overfitting. The calculation of $L2$ regularisation is as follows, where L_{reg} represents the $L2$ regularisation term, λ represents the weight decay factor used to control the strength of regularisation, and w_i represents the weighting parameters in the model.

$$L_{reg} = \lambda \sum_{i=1}^{n} w_i^2 \tag{6}$$

Combining the aforementioned improved algorithms ESMOTE and CSGAN, this paper proposes an efficient balanced process for railway infrastructure signal data. The initial signal data is input into the ESMOTE method, where relevant parameters are set to conduct experiments generating one-dimensional vibration data in different states. Meanwhile, the time-frequency diagram data of railway facility signals is fed into the CSGAN model through multiple rounds of training and the best model is saved. Subsequently, the generator part is extracted from the optimal model, trained weights are loaded, and used to generate time-frequency diagrams for different types of conditions, providing rich and high-quality data support for railway infrastructure fault diagnosis tasks.

4 Fault diagnosis system for railway infrastructure based on multi-scale ResNet and transfer learning

4.1 Multi-scale residual convolution

To address the issue where traditional deep learning needs substantial data for model training during early stages of railway infrastructure fault diagnosis, and some operating condition data is difficult to obtain in large quantities, this paper proposes a transfer learning model with multi-scale residual feature extraction. The model adopts an improved ResNet as its core architecture. Three channels are constructed, utilising different receptive field sizes so the model can extract railway infrastructure fault signal features at various scales, then fuses these features through weighted integration to enhance the network's characteristic extraction capability. Ultimately, a subdomain adaptation-based transfer learning strategy is applied, enabling railway infrastructure fault diagnosis under complex operating conditions.

In railway infrastructure fault diagnosis under different operational conditions, changes in the operational state of facilities present greater challenges for diagnosis. This paper suggests a multi-scale residual convolutional neural network architecture that addresses insufficient feature extraction capability and deep network degradation issues in traditional networks under variable operational conditions through multi-scale feature fusion and ResNet mechanisms. The network adopts a three-level multi-scale feature extraction module, with each scale module containing three residual block structures configured with 1×3, 1×5, and 1×7 convolution kernels of various scales to capture fault signal characteristics from railway infrastructure components of various sizes, producing 64, 128, and 256 channel outputs.

The original signal undergoes preliminary feature extraction using a 1×7 convolution kernel to avoid excessive parameters within the model input, followed by data base features obtained through 2×2 max pooling. The base features are then separately fed into three multi-scale processing modules where different scale ResNet deeply mine data features of various scales. Finally, each submodule outputs a 1 × 256 feature that undergoes average pooling to obtain a fused output feature of size 1 × 768.

4.2 Domain adaptation

In addressing data distribution shifts across different operational condition railway infrastructure fault diagnosis, global domain adaptation maps source and target domain data into a common characteristic space via feature space mapping to align marginal distributions. Subdomain adaptation differs from global domain adaptation by precisely aligning conditional probability distributions of each subdomain in the source and target domains, thereby gradually bringing their marginal distributions closer during transfer learning. This paper adopts the local maximum mean discrepancy (LMMD) (Huang et al., 2023) as the measurement of subdomain distribution differences to construct an optimisation objective for conditional distributions. The distribution difference between source domain features $X_s^l = \{x_i^{s,l}\}_{i=1}^{n_s}$ and target domain features $X_t^l = \{x_j^{s,l}\}_{j=1}^{n_t}$ is shown in equation (7).

$$d^2(p,q) = \frac{1}{C}\sum_{c=1}^{C}\left\| \sum_{x_i^s \in D_s} w_i^{sc}\phi\left(x_i^s\right) - \sum_{x_i^t \in D_t} w_i^{tc}\phi\left(x_j^t\right)\right\|_H^2 \tag{7}$$

where w_i^{sc} and w_i^{tc} are the weight values of x_i^s and x_j^t for fault category *c*, individually. $\sum_{i=1}^{n_s} w_i^{sc} = 1$ and $\sum_{i=1}^{n_s} w_i^{sc} = 1$ are the weighted sums of samples in class *c* after mapping. The weight w_i^c is shown in equation (8), where y_j^c is the c^{th} element of feature y_i.

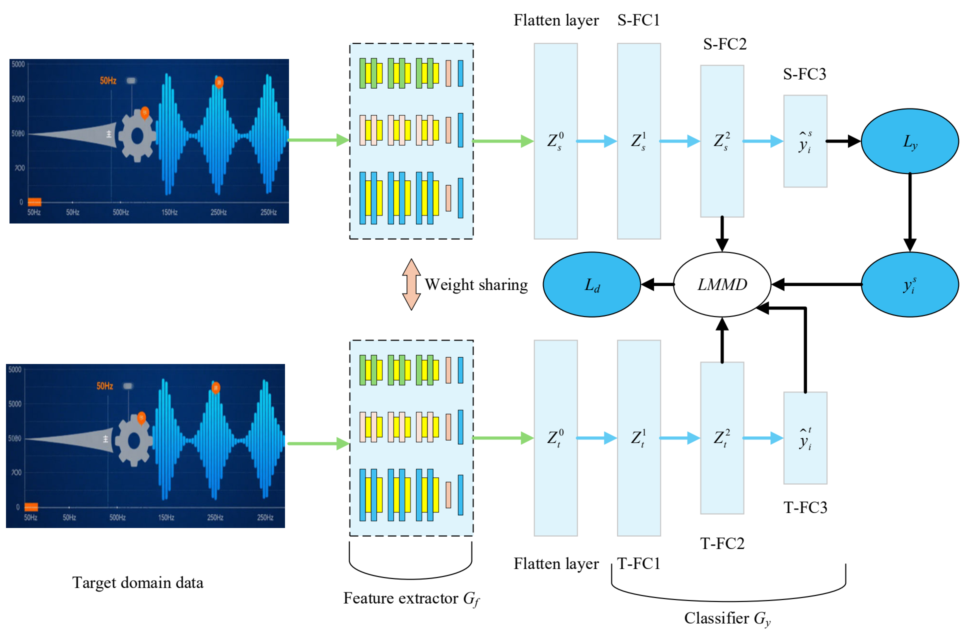
$$w_i^c = \frac{y_i^c}{\sum_{(x_j, y_j) \in D} y_j^c} \tag{8}$$

In the source domain dataset, sample weights w_i^{sc} are directly calculated using equation (8) based on the label vector y_i^s of source domain data y_i. However, since there are no labelled vectors for target domain data, it is not possible to determine the weight through this method. This paper proposes utilising the probability distribution output by deep neural networks as a pseudo-label estimation approach for target domain samples. Inputting the target domain sample x_i^t into the trained diagnosis network, use the $\hat{y}_i^t$ output from the softmax layer to represent the probability of the sample belonging to fault types.

4.3 Railway infrastructure fault diagnosis system model

In actual fault diagnosis processes, the amount of railway infrastructure fault data under certain operating conditions is tough to satisfy the needs for training models using traditional deep learning methods. When introducing transfer learning algorithms, changes in the operating conditions of railway infrastructure lead to distribution differences in diagnostic data, and obtaining labels for samples to be diagnosed also presents some difficulties. To cope with these issues, this chapter puts forward a railway infrastructure fault diagnosis system in light of multi-scale ResNet and transfer learning, intended for railway infrastructure fault diagnosis under complex operating conditions. The entire structure of the suggested railway infrastructure fault diagnosis system is implied in Figure 1. By characteristic space mapping, labelled source domain data and unlabelled target domain data are uniformly transferred to a subspace. This approach effectively utilises the common characteristics exhibited by fault signals of railway infrastructure under different operating conditions. LMMD is adopted as the subdomain adaptation metric to perform subdomain adaptation on the extracted multi-condition shared features, achieving multiple condition fault diagnosis for railway infrastructure.

Figure 1 The entire structure of the suggested railway infrastructure fault diagnosis system (see online version for colours)



The proposed model consists primarily of two parts: a data characteristic extractor and a classifier. The above-mentioned multi-scale ResNet serves as the data feature extractor G_f of this model, while the classifier is labelled G_y. An adaptive layer selects the FC2 layer, as shown in Figure 1, to reduce the linear maximum mean deviation between source field and target field through the adaptive layer, thus reducing data distribution differences caused by changes in operating conditions.

When the source field $D_s = \{(x_i^s, y_i^s)\}_{i=1}^{n_s}$ and target field $D_t = \{(x_j^t)\}_{j=1}^{n_t}$ data are input into the feature extraction model, multi-scale features will be extracted. These features capture detailed information at different levels of the data. After passing through a flattening layer, these features result in features $Z_s^0 = \{z_i^{s,0}\}_{i=1}^{n_s}$ and $Z_t^0 = \{z_j^{t,0}\}_{j=1}^{n_t}$, while features Z_s^0 and Z_t^0 are adapted by the subdomain adaptation layer of classifier G_y, resulting in adaptive features $Z_s^2 = \{z_i^{s,2}\}_{i=1}^{n_s}$ and $Z_t^2 = \{z_j^{t,2}\}_{j=1}^{n_t}$. Combined with label vector y_i^s and predicted label vector $\hat{y}_i^t$, the source field and target field subdomain adaptation loss values L_d are obtained.

$$L_d = L\hat{MMD}_H^2 (Z_s^2, Z_t^2) \tag{9}$$

The classification output of the source field data through the classifier is $\hat{y}_i^s$. The classification loss value for the source domain L_y is calculated using a normal source domain label vector, as described below, where $J(\cdot,\cdot)$ represents the cross-entropy loss function.

$$L_y = \frac{1}{n_s} \sum_{i=1}^{n_s} J\left(\hat{y}_i^s, y_i^s\right) \tag{10}$$

In the backpropagation process of the proposed model, only the classification loss L_y is optimised with respect to a standard optimisation function. In this method, both the classification loss L_y and adaptation loss L_d must be simultaneously optimised. The optimisation function formula is as follows, where $\theta = \{w, b\}$ represents the bias and weights of the model; λ is the adaptive weight parameter, $0 < \lambda < 1$.

$$\min_{\theta=\{w,b\}} \left(L_y + \lambda L_d\right) \tag{11}$$

$$\lambda = \frac{2}{1 + e^{-10t/T}} - 1 \tag{12}$$

where t is the current iteration count during training, and T is the total number of iterations.

The final railway infrastructure fault diagnosis model achieves minimum linear LMMD values, enabling fault identification and classification under different operating conditions, thereby improving diagnostic generalisation performance.

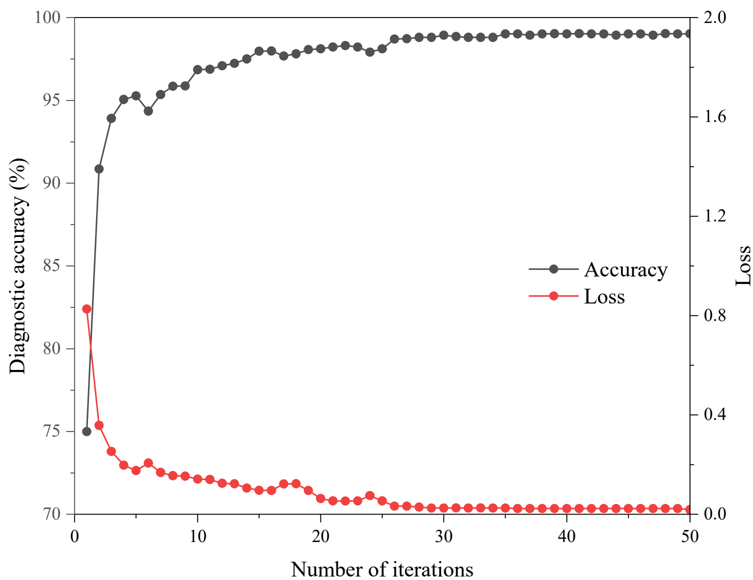
5 Analysis of experimental results

5.1 Railway infrastructure fault diagnosis results analysis

This paper adopts bearing data from the bearing test bench at Shijiazhuang Tiedao University. The experiment uses double-row tapered roller bearings. There are eight types of bearing health states: normal, outer ring minor fault, outer ring severe fault, inner ring minor fault, inner ring severe fault, roller severe fault, roller minor fault, and compound faults between the outer ring and rollers. These eight fault types are denoted as T1, T2,..., T8. All these faults represent real faults occurring during railway infrastructure operations.

To make the experimental test more accurate, after balancing the fault data using the ESMOTE-CSGAN method proposed in this paper, 1,000 sets of signals are chosen for each type of bearing fault, totalling 8,000 sets. Each signal contains 1,024 points. For each type of faulty bearing, 750 samples are selected as the training set and 250 as the test set. Training samples and validation samples in the training set are randomly divided at a ratio of 4:1, and the model is trained for 50 rounds to obtain the optimal model. Then, the test set is input into the optimal model for testing. The computer hardware configuration includes an AMD R7-3700X processor, 32G RAM, and an RTX3070 GPU. The deep learning network framework was developed using PyTorch with Python 3.8 as the programming language. The batch size is set to 64, the studying rate to 0.01, the optimiser's momentum value to 0.9, and the weight decay to 0.0004. The amount of training iterations is set to 300.

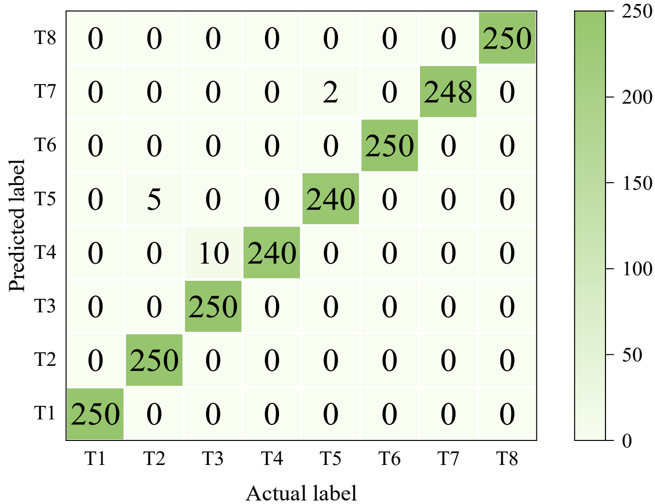
Figure 2 Training results of the MSRN-TL model (see online version for colours)



The time-frequency images generated by the ESMOTE-CSGAN data balancing method for various fault states are input into the model. After 50 iterations, the proposed MSRN-TL model obtains results as implied in Figure 2. The training accuracy of the suggested model reaches 98.99%. After 30 iterations, the model loss stabilises and becomes smooth. This verifies that the model achieves good performance in railway infrastructure fault detection.

To more intuitively demonstrate the model's performance in railway infrastructure fault detection, a confusion matrix is introduced for visual analysis of the test results, as shown in Figure 3. Out of 2,000 test samples, 22 are misclassified: 10 samples from T4 are classified into T3, with an accuracy rate of 96% for T4. In T5, 5 samples are classified into T6 and another 5 into T2, achieving a recognition accuracy of 96%. For T7, two samples are misclassified into T5, resulting in an accuracy rate of 99.2%. The model's overall recognition accuracy is 98.9%, recall is 98.9%, precision is 98.93%, and the F1 score is 98.89%. This verifies that the MSRN-TL model achieves excellent performance in railway infrastructure fault detection.

Figure 3 Visualisation of fault detection for the proposed MSRN-TL model (see online version for colours)



5.2 Diagnostic performance comparison

To further verify the fault diagnostic performance of the proposed model, this paper selects RN-Trans [15], DFP-TL [16], and Conv-GAN [18] as baseline models. Evaluation metrics include accuracy, precision, F1 score, and specificity. Figure 4 shows a comparison of loss and accuracy during training for the four models. After 40 iterations, MSRN-TL and Conv-GAN achieve diagnostic accuracies of 92.08% and 91.56%, respectively. RN-Trans converges after 80 iterations with an accuracy of 90.07%. DFP-TL converges after 55 iterations, achieving a precision of 91.55%. The MSRN-TL model not only demonstrates high diagnostic accuracy but also achieves low loss, performing well in railway infrastructure fault diagnosis tasks.

Figure 4 A comparison of loss and accuracy during training for the four models, (a) loss function (b) diagnostic accuracy (see online version for colours)

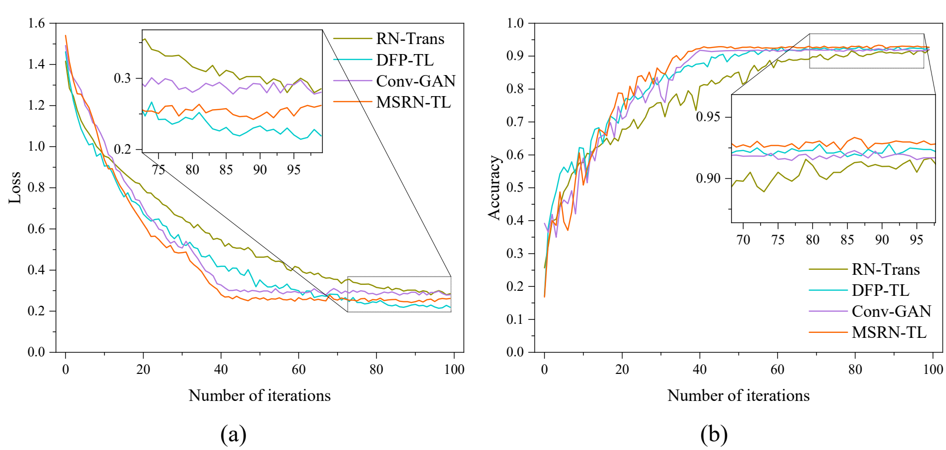


Table 1 Fault diagnosis performance comparison

Model	*Accuracy/%*	*F1/%*	*Specificity/%*
RN-Trans	90.46	89.01	95.39
DFP-TL	91.51	91.97	96.05
Conv-GAN	92.59	92.81	97.05
MSRN-TL	93.86	94.28	98.47

Table 1 presents the comparison of various diagnostic performance metrics for each model. The Accuracy of MSRN-TL is 93.86%, which represents an improvement of 3.4%, 2.35%, and 1.27% over RN-Trans, DFP-TL, and Conv-GAN, respectively. Comparing F1 and Specificity further, the improvements for MSRN-TL are 5.27% and 3.08% against RN-Trans, 2.31% and 2.42% against DFP-TL, and 1.47% and 1.42% against Conv-GAN.RN-Trans mainly combines deep ResNet with transformer for railway infrastructure fault diagnosis; however, under limited data conditions, RN-Trans tends to memorise noise and specific patterns in the training set rather than learning generalised fault features, resulting in reduced generalisation ability in real-world scenarios. DFP-TL utilises deep feature representation combined with transfer learning methods for railway infrastructure fault diagnosis. However, the types of faults contained in source domain data might not fully align with those actually occurring in target domain railway infrastructure. The model may fail to recognise fault patterns it has never encountered before. Conv-GAN addresses dataset imbalance through GAN and achieves fault feature extraction and classification via CNN, thereby improving diagnostic accuracy to some extent; however, the complexity of CNN-based feature extraction is relatively high. MSRN-TL can effectively capture common features of fault patterns under different working conditions of railway infrastructure and deeply mine the intrinsic consistency of fault features through multi-scale ResNet, enhancing the model's capability for feature learning.

6 Conclusions

To address the issue of scarce railway infrastructure fault data and low diagnostic accuracy caused by imbalanced datasets in existing research, this paper presents an intelligent fault diagnosis system for railway infrastructure that leverages deep learning. First, to tackle signal dataset imbalance issues specific to railway infrastructure, a data balancing method combining ESMOTE with CSGAN is proposed, leveraging the characteristics of the data itself. An unsupervised clustering algorithm and natural neighbour concept are introduced to improve sample generation efficiency and enhance the quality of generated samples. The improved data balancing approach increases training stability by incorporating additional class label information and optimising the training loss function, thereby generating higher-quality fault samples. Based on this, transfer learning methods are introduced to migrate knowledge learned from source domain data to the target domain, enabling fault identification under complex working conditions. By constructing a feature extractor combining ResNet with multi-scale feature fusion technology, the network's feature extraction capability is enhanced while weakening the impact of operating condition changes on diagnostic results, and effectively addressing the issue of network degradation as network depth increases. Then, a subdomain adaptation transfer learning strategy is adopted to achieve fault diagnosis for railway infrastructure. Experimental outcome implies that the suggested approach achieves a diagnostic accuracy of 93.86%, which is at least 1.27% higher than baseline models and can accurately realise fault diagnosis for railway infrastructure.

Acknowledgements

This work is supported by the Foundation of Major science and technology projects in Sichuan Province (No. 2023ZDZX0010), Foundation of China State Railway Group Company Limited (No. 2024YJ178), Foundation of China State Railway Group Company Limited (No. Q2024G018), Foundation of China Academy of Railway Sciences Corporation Limited (No.2024YJ311) and Foundation of Scientific research project of Beijing Huatie Information Technology Limited (No. 2024HT06).

Declarations

All authors declare that they have no conflicts of interest.

References

Abu-Srhan, A., Abushariah, M.A. and Al-Kadi, O.S. (2022) 'The effect of loss function on conditional generative adversarial networks', *Journal of King Saud University-Computer and Information Sciences*, Vol. 34, No. 9, pp.6977–6988.

Ahmed, M., Seraj, R. and Islam, S.M.S. (2020) 'The k-means algorithm: a comprehensive survey and performance evaluation', *Electronics*, Vol. 9, No. 8, pp.12–20.

Chen, Y., Song, B., Zeng, Y., Du, X. and Guizani, M. (2021) 'Fault diagnosis based on deep learning for current-carrying ring of catenary system in sustainable railway transportation', *Applied Soft Computing*, Vol. 100, pp.69–77.

Duan, F., Zhang, S., Yan, Y. and Cai, Z. (2022) 'An oversampling method of unbalanced data for mechanical fault diagnosis based on MeanRadius-SMOTE', *Sensors*, Vol. 22, No. 14, pp.51–66.

Elreedy, D., Atiya, A.F. and Kamalov, F. (2024) 'A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning', *Machine Learning*, Vol. 113, No. 7, pp.4903–4923.

Fang, W., Yu, Z., Chen, Y., Huang, T., Masquelier, T. and Tian, Y. (2021) 'Deep residual learning in spiking neural networks', *Advances in Neural Information Processing Systems*, Vol. 34, pp.21056–21069.

Ghosh, C., Verma, A. and Verma, P. (2022) 'Real time fault detection in railway tracks using fast Fourier transformation and discrete wavelet transformation', *International Journal of Information Technology*, Vol. 14, No. 1, pp.31–40.

Han, X. (2024) 'Fault diagnosis model for railway signalling equipment using deep learning techniques', *International Journal of Sensor Networks*, Vol. 45, No. 1, pp.40–53.

Hu, H., Tang, B., Gong, X., Wei, W. and Wang, H. (2017) 'Intelligent fault diagnosis of the high-speed train with big data based on deep neural networks', *IEEE Transactions on Industrial Informatics*, Vol. 13, No. 4, pp.2106–2116.

Hu, X., Cao, Y., Tang, T. and Sun, Y. (2022) 'Data-driven technology of fault diagnosis in railway point machines: Review and challenges', *Transportation Safety and Environment*, Vol. 4, No. 4, pp.36–41.

Huang, M., Yin, J., Yan, S. and Xue, P. (2023) 'A fault diagnosis method of bearings based on deep transfer learning', *Simulation Modelling Practice and Theory*, Vol. 122, pp.51–59.

Huang, Y., Lin, J., Liu, Z. and Wu, W. (2019) 'A modified scale-space guiding variational mode decomposition for high-speed railway bearing fault diagnosis', *Journal of Sound and Vibration*, Vol. 444, pp.216–234.

Lv, J., Xiao, Q., Zhai, X. and Shi, W. (2024) 'A high-performance rolling bearing fault diagnosis method based on adaptive feature mode decomposition and transformer', *Applied Acoustics*, Vol. 224, pp.16–23.

Men, Z., Li, Y., Gao, L. and Zhang, Z. (2025) 'Fault diagnosis method for railway wagon bearings under imbalanced dataset based on improved ACWGAN', *Nonlinear Dynamics*, Vol. 113, No. 12, pp.14935–14962.

Raza, A., Sehar, R., Moiz, A., Alluhaidan, A.S., El-Rahman, S.A. and AbdElminaam, D.S. (2025) 'Novel conditional tabular generative adversarial network based image augmentation for railway track fault detection', *PeerJ Computer Science*, Vol. 11, pp.28–34.

Shao, S., McAleer, S., Yan, R. and Baldi, P. (2018) 'Highly accurate machine fault diagnosis using deep transfer learning', *IEEE Transactions on Industrial Informatics*, Vol. 15, No. 4, pp.2446–2455.

Shao, Y., Yuan, X., Zhang, C., Song, Y. and Xu, Q. (2020) 'A novel fault diagnosis algorithm for rolling bearings based on one-dimensional convolutional neural network and INPSO-SVM', *Applied Sciences*, Vol. 10, No. 12, pp.43–50.

Simone, L., Caputo, E., Cinque, M., Galli, A., Moscato, V., Russo, S., Cesaro, G., Criscuolo, V. and Giannini, G. (2023) 'LSTM-based failure prediction for railway rolling stock equipment', *Expert Systems with Applications*, Vol. 222, pp.42–53.

Song, Y., Huang, L., Xu, P. and Wan, M. (2021) 'Real-time rail fault diagnosis based on vibration signal analysis and second-order sinusoidal model', *IEEE Sensors Journal*, Vol. 22, No. 4, pp.3382–3396.

Sun, Y., Cao, Y. and Li, P. (2022) 'Contactless fault diagnosis for railway point machines based on multi-scale fractional wavelet packet energy entropy and synchronous optimization strategy', *IEEE Transactions on Vehicular Technology*, Vol. 71, No. 6, pp.5906–5914.

Sun, Y., Cao, Y., Li, P. and Su, S. (2023) 'Entropy feature fusion-based diagnosis for railway point machines using vibration signals based on kernel principal component analysis and support vector machine', *IEEE Intelligent Transportation Systems Magazine*, Vol. 15, No. 6, pp.96–108.

Wang, Q., Chen, S., Zeng, J., Du, W. and Wei, L. (2024) 'A deep learning fault diagnosis method for metro on-board detection on rail corrugation', *Engineering Failure Analysis*, Vol. 164, pp.86–93.

Wang, X., Jiang, H., Mu, M. and Dong, Y. (2025) 'A trackable multi-domain collaborative generative adversarial network for rotating machinery fault diagnosis', *Mechanical Systems and Signal Processing*, Vol. 224, pp.23–29.

Wu, X., Yang, W. and Cao, J. (2023) 'Fault diagnosis of railway point machines based on wavelet transform and artificial immune algorithm', *Transportation Safety and Environment*, Vol. 5, No. 4, pp.72–84.

Wu, Z., Shen, C. and Van Den Hengel, A. (2019) 'Wider or deeper: Revisiting the ResNet model for visual recognition', *Pattern Recognition*, Vol. 90, pp.119–133.

Zhang, Y., Xu, T., Chen, C., Wang, G., Zhang, Z. and Xiao, T. (2021) 'A hierarchical method based on improved deep forest and case-based reasoning for railway turnout fault diagnosis', *Engineering Failure Analysis*, Vol. 127, pp.54–67.

Zhu, Q., Feng, J. and Huang, J. (2016) 'Natural neighbor: A self-adaptive neighborhood method without parameter K', *Pattern Recognition Letters*, Vol. 80, pp.30–36.