# Intelligent classification of ballast bed defects using a bimodal deep learning model

Bu, Junjie; Jing, Guoqing; Long, Xujie; Wang, Lei; Peng, Zhan; Guo, Yunlong

**DOI**
10.1016/j.trgeo.2024.101464

**Publication date**
2025

**Document Version**
Final published version

**Published in**
Transportation Geotechnics

Transportation Geotechnics 50 (2025) 101464

Contents lists available at ScienceDirect

Transportation Geotechnics

journal homepage: www.elsevier.com/locate/trgeo

Original Article

# Intelligent classification of ballast bed defects using a bimodal deep learning model

Junjie Bu [a], Guoqing Jing [a], Xujie Long [b], Lei Wang [c], Zhan Peng [a,d,*], Yunlong Guo [e]

[a] School of Civil Engineering, Beijing Jiaotong University, Beijing 100044, China
[b] School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China
[c] Department of Civil and Architectural Engineering and Construction Management, University of Cincinnati, Cincinnati, OH 45221 USA
[d] Infrastructure Inspection Research Institute, China Academy of Railway Sciences Corporation Limited, Beijing 100081, China
[e] Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft 2628CN, Netherlands

ARTICLE INFO

*Keywords:*
Ballast Bed Defects Detection
GPR
Deep Learning
Model Fusion

ABSTRACT

The method of detecting ballast bed defects using ground penetrating radar (GPR) is an important method for guiding the maintenance of railway infrastructure. Currently, this technology primarily relies on time–frequency analysis to assess the condition of the ballast bed and manual interpretation of GPR images to identify defect areas and types, resulting in low automation levels. This paper proposes a bimodal deep learning classification model that enables intelligent classification of moisture and mud pumping defects in ballast beds. This model includes two channels, each processing a different data modality. One channel uses a Multilayer Perceptron (MLP) to extract features of A-scan data in the time domain. The other channel utilizes Short-Time Fourier Transform (STFT) to convert time domain signals into frequency domain signals, which are then processed by a ResNet18 to extract frequency domain features. By fusing the time and frequency features, the proposed Time-Frequency-Fusion ResNet model (TFF-ResNet) demonstrates superior performance. Experimental results show that TFF-ResNet outperforms the standalone MLP and ResNet18 models, with performance improvements of approximately 24% and 14% on the validation dataset, and 21% and 34% on the testing dataset, respectively.

## Introduction

The ballast bed, as a crucial component of ballasted tracks, has the function of supporting the track structure, providing longitudinal and lateral resistance, and providing drainage [1]. However, with increased intensive high-speed train loading, axle load, and traffic volume, the condition of the ballast bed and subgrade deteriorates, leading to numerous defects. Fouling in the ballast bed primarily arises from fines due to ballast abrasion and breakage, infiltration from external sources, and degradation of other components [2]. The stiffness of the ballast bed increases, and resilience reduces when fouling jams the voids between ballast particles. Under the stress from the train, ballast particles and subgrade begin to mix at the interface, resulting in the formation of ballast pockets. The ballast pocket usually stores a lot of water, which causes the subgrade soil to become muddy[3]. Mud pumping is a defect characterized by water mixed with fouling being drawn up to the subsurface of the ballast layer. Extensive engineering practice has demonstrated that mud pumping is one of the most common defects in railway ballast beds, significantly impacting train safety and necessitating timely maintenance [4,5,6,7]. Due to the limited developments in track inspection equipment and technology, the current ballast bed condition is mainly assessed by indirect means through the total passing gross load. But relying only on this metric is unreasonable and can sometimes lead to incorrect maintenance decisions. The GPR method for ballast bed inspection is characterized by its high speed and nondestructive nature, making it an important technological approach for railway infrastructure monitoring [8,9]. However, traditional GPR data analysis methods mainly rely on preset parameters to filter noise, extract signal features in the time–frequency domain, and then judge the condition of the ballast bed according to these features. [10,11,12,13,14]. This process requires multiple adjustments according to different geological conditions and ballast bed structures and is highly dependent on expert experience and manual intervention, leading to low efficiency and poor adaptability. As artificial intelligence progresses, the application of machine learning and deep learning is transforming traditional GPR data analysis methods. By automating the optimization of processing parameters and

* Corresponding author at: School of Civil Engineering, Beijing Jiaotong University, Beijing 100044, China.
*E-mail address:* 23115139@bjtu.edu.cn (Z. Peng).

https://doi.org/10.1016/j.trgeo.2024.101464
Received 5 May 2024; Received in revised form 28 October 2024; Accepted 8 December 2024
Available online 10 December 2024

learning from large datasets, these algorithms not only enhance the precision and efficiency of detection but also significantly reduce reliance on manual operations and improve adaptability to various complex conditions. These algorithms can be divided into two main categories based on the nature of their learning targets: machine learning algorithms based on the single-channel waveform (A-scan data) and deep learning algorithms based on the two-dimensional profile (B-scan data).

(1) Machine learning algorithms based on the A-scan data. This method primarily consists of three steps: data preprocessing, feature extraction, and classification using algorithms. The goal of data preprocessing is to enhance the signal-to-noise ratio of the data. Typical preprocessing methods include background denoising, zero offset removal, and band-pass filtering [15,16]. The feature extraction process involves selecting representative features of defects based on theoretical knowledge. In the time domain, this includes peak, mean, variance, skewness, kurtosis, and fourth-order moments of the A-scan data, while in the frequency domain, it includes entropy of the frequency spectrum and energy density spectrum [17,18]. These features should be invariant to time shifts and scaling, insensitive to noise and multiple reflections, and easily discriminable [19]. Classification algorithms learn from extracted features to enable intelligent recognition of data, with the main algorithms currently including the K-nearest neighbors algorithm (KNN), SVM, neural networks, etc. Sezgin et al. [20] applied a KNN algorithm in conjunction with PCA to A-scan data for identifying underground targets. While the KNN algorithm is effective at differentiating between various feature classes, its accuracy is notably reduced by the sensitivity to the parameter K, especially when the data set contains outliers [21]. Shao et al. [22] introduced an automatic classification system that assesses ballast bed conditions by extracting magnitude spectra at salient frequencies and classifying them using SVM. Du et al. [23] extracted segmented energy, variance, and interface as the eigenvalues from GPR data to establish a neural network model for the identification of mud-pumping in railway subgrade. Such methods require manual feature extraction from GPR data, and the accuracy of these algorithms highly depends on the appropriateness of the feature extraction. Therefore, these algorithms only perform well on specific datasets. Deep learning offers robust capabilities for automatic feature extraction. Xu [24] and Ahmadvand et al. [25] both used 1D-CNN to extract features from A-scan signals to establish a model for identifying defects in concrete pavements. Liu et al. [26] combined the 1D-CNN and RNN models to extract features from A-scan signals for assessing the condition of railway subgrade. However, they only extract A-scan data features from the time domain, and the scene generalization ability of the 1D-CNNorithm is limited.
(2) Deep learning algorithms based on the B-scan data. CNN and RNN are the most popular deep learning architectures used in the application of GPR B-scan data. Özkaya et al. [27] combined residual CNN and Bi-LSTM models for analyzing B-scan data. The method shows high performance in determining the scanning frequency of GPR B-Scan data and the type of soil. Xu et al. [28] have enhanced recognition accuracy by integrating improvement strategies such as feature cascading, the adversarial spatial dropout network (ASDN), Soft-NMS, and data augmentation into the Faster R-CNN framework, tailored to the characteristics of subgrade defects. MA et al. [29] proposed an LS-YOLOv3 model for real-time detection of railway subgrade defects, utilizing a deep residual network to extract characteristic features. Lei et al. [30] introduced an adaptive target region detection algorithm and a combined Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM) framework, achieving remarkable accuracy in classifying hyperbolic signatures. Although deep learning algorithms can effectively extract features from grayscale images, using grayscale images as training data presents the following problems: (1) Since GPR data is stored as 16-bit integers, ranging from 0 to 65,535 (2ˆ16——1), converting it into 8-bit grayscale images results in a reduced range from 0 to 255. This compression of the dynamic range consequently diminishes detail, particularly in areas with subtle features. (2) When creating grayscale images, the data from the entire railway line needs to be segmented into uniform sizes. This segmentation method may disrupt the integrity of the defect areas, making the morphology of the defects more complex and thus increasing the difficulty of model training. (3) Each image may contain defects of varying numbers, sizes, and types, and the boundaries of these defect areas are often blurred. This can lead to errors such as omissions, incorrect selections, and irregular bounding boxes during the annotation process. (4) In actual ballast bed inspections, the method described in this paper can quickly classify data along the entire line and locate defects in real-time.

The problems of the above two methods include the need to manually design features and poor model generalization, and the production of image datasets is laborious and does not contain actual physical information. Considering the limitations of the above methods, this paper proposes a deep learning algorithm based on A-scan data that enables intelligent classification of moisture and mud pumping defects in ballast beds. This study extracts features from both the time domain and frequency domain simultaneously to gather physical information from A-scan signals. Classical time domain feature processing algorithms, such as LSTM and GRU, designed to capture long-term dependencies, are not suitable for A-scan samples with short data lengths. The MLP focuses on minimizing output error, allowing it to accurately learn key data characteristics even with limited samples. Additionally, the MLP, with its simple network structure, has high computational efficiency, making it suitable for real-time or near-real-time GPR data processing [31]. Therefore, considering the GPR data characteristics and practical detection requirements, choosing the MLP to extract time domain features is more reasonable. To extract frequency domain features while retaining time information, this study employs the commonly used STFT algorithm, converting A-scan data into two-dimensional spectral images that contain both time and frequency information [32]. By analyzing these spectral images, the frequency domain characteristics of A-scan data are effectively extracted. GPR spectral images typically exhibit highly structured features that differ from natural images and display complex patterns in both time and frequency dimensions. Resnet18 leverages deep residual blocks, allowing the network to efficiently learn deep features without losing signal strength, which is crucial for analyzing subtle frequency variations in GPR data. Moreover, compared to other classic image feature extraction algorithms such as AlexNet and VGG, Resnet18 has a more efficient and simplified architecture, which speeds up the training process and allows for better convergence. Therefore, this paper selects Resnet18 to extract the frequency domain features of the A-scan signal.

## Data preparation

### *Principle of GPR*

GPR is a geophysical survey method that uses electromagnetic radiation to image the subsurface. It comprises a transmitting antenna, a receiving antenna, and a main control unit. The transmitting antenna emits electromagnetic waves with a fixed bandwidth into the ballast layer. When the waves encounter targets with different dielectric properties, the waves are reflected and scattered. Then, the receiving antenna receives the return waves, and the main control unit samples and stores the electrical signals, obtaining the A-scan data of the current measurement point. When using GPR for railway ballast bed inspection, the antenna moves along the railway line, and the A-scan data collected at each point forms a two-dimensional radar image. The propagation of electromagnetic waves in a medium follows Maxwell's equations, as shown in Eq. (1) [33]:

$$\begin{cases} \nabla \times H = J + \dfrac{\partial D}{\partial t} \\ \nabla \times E = -\dfrac{\partial B}{\partial t} \\ \nabla \bullet D = \rho \\ \nabla \bullet B = 0 \end{cases} \tag{1}$$

where $H(A/m)$ is the magnetic field strength, $E(V/m)$ is the electric field strength, $D(C/m^2)$ is the electric flux density, $B(Wb/m^2)$ is the magnetic flux density, $J(A/m^2)$ is the current density, and $\rho(C/m^3)$ is the charge density.

In the Chinese railway ballast standards, the ballast typically consists of non-magnetic materials such as granite and basalt. Therefore, this paper focuses on the influence of permittivity on electromagnetic wave propagation.

When electromagnetic waves encounter an interface with a significant difference in permittivity (such as ballast and air), reflection occurs. In physics, the strength of reflection is defined by the reflection coefficient, which can be calculated using Eq. (2) [34]:

$$r_{j,j+1} = \frac{\sqrt{\varepsilon_j} - \sqrt{\varepsilon_{j+1}}}{\sqrt{\varepsilon_j} + \sqrt{\varepsilon_{j+1}}} \tag{2}$$

where $r_{j,j+1}$ is the reflection coefficient, $\varepsilon_j$ is the relative permittivity of the material above the interface, $\varepsilon_{j+1}$ is the relative permittivity of the material below the interface. The greater the difference between $\varepsilon_j$ and $\varepsilon_{j+1}$, the greater the absolute value of the $r_{j,j+1}$, indicating more energy is reflected and the waveform amplitude is larger.

Additionally, electromagnetic waves attenuate as they propagate, and physics defines the attenuation coefficient to measure the medium's effect on the attenuation of electromagnetic waves, as shown in Eq. (3) [34]:

$$\alpha = \omega\sqrt{\frac{\mu\varepsilon}{2}\left[\sqrt{1+\left(\frac{\sigma}{\omega\varepsilon}\right)^2} - 1\right]} \tag{3}$$

where $\alpha$ is the attenuation constant, and $\omega$ is the angular frequency. For non-magnetic materials like railway ballast, $\mu$ (magnetic permeability) does not vary significantly. The permittivity $\varepsilon$ and electrical conductivity $\sigma$ are the main factors influencing the attenuation coefficient.

Due to the differences in $\varepsilon$ and $\sigma$ between air, ballast, water, and mud, there are variations in the scattering, reflection, and attenuation of electromagnetic waves in these media. Based on these differences, deep learning algorithms can be utilized to intelligently classify different ballast bed conditions.

### Data collection

During the detection process, a total of three survey lines are arranged: one at the center of the railway and the other two on both sides of the rails, as depicted in Fig. 1(a). This experiment primarily involves collecting GPR data from the ballast bed and subgrade, with the data acquisition parameters detailed in Table 1. To expand the geographic coverage of the dataset and enhance the model's generalization ability, GPR data were collected from four general-speed railways in North China, Central China, and Southwest China. The foundation for defect determination in this study is twofold: expert interpretation based on the GPR characteristics of ballast bed defects, and on-site excavation verification. The collected ballast samples were washed, dried, and sieved to measure the water content and fouling level of each railway section, as illustrated in Fig. 1(b). In different railway lines and sections, the distribution of moisture anomalies and mud-pumping defects varies considerably. Moisture anomalies primarily occur near the subballast layer, while mud-pumping defects are observed across the entire cross-section of the ballast bed. The types and quantities of A-scan data collected from each line are detailed in Table 2.

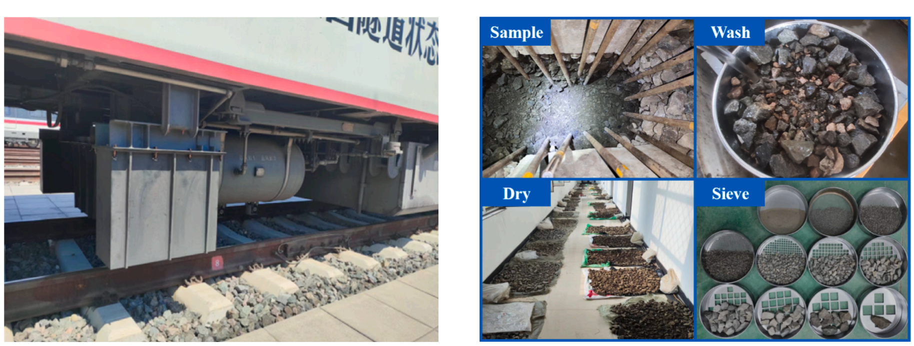
### Data preprocessing

Due to interference from external environments, the original data exhibits a low signal-to-noise ratio. Processing this data is crucial for suppressing noise and enhancing the visibility of key features, which include interlayer interfaces and the amplitude, phase, and waveform of reflected waves from target defects.

1. Background denoising. The sources of noise primarily come from two parts: electromagnetic interference noise and standing wave interference. Electromagnetic interference noise mainly originates from the disruptive signals emitted by train operation control equipment

**Table 1**
Main technical parameters set for the operation of GPR equipment.

| Parameter | Values |
|---|---|
| Sampling interval | 12 cm |
| Detection Depth | 200 cm |
| Center Frequency | 400 MHz |
| Sampling Frequency | 7.8 GHz |
| Time Window | 65 ns |
| Number of Sampling Points | 512 |




(a) vehicle-mounted GPR

(b) determine the kinds of ballast bed defects through on-site excavation.

**Fig. 1.** Collection of GPR data and qualitative assessment of ballast bed defects.

**Table 2**
Number of A-scan data obtained on each railway line.

| Line | Normal | Moisture | Mud pumping | Total |
|---|---|---|---|---|
| A | 5900 | 5800 | 5800 | 17,500 |
| B | 5900 | 5700 | 0 | 11,600 |
| C | 5900 | 5800 | 5800 | 17,500 |
| D | 3852 | 2948 | 2520 | 9320 |

and communication devices. Typically, these noises appear randomly and differ in frequency from the main frequency of the GPR transmission wave, so they can be removed through band-pass filtering. In this study, the central frequency of the electromagnetic wave is set at 400 MHz. The band-pass filtering involves removing low-frequency interferences below 100 MHz and high-frequency interferences above 800 MHz. By doing so, the filter effectively reduces noise that significantly deviates from the central frequency, thereby enhancing the quality of the GPR data and the reliability of the detection results. Another source of noise is standing wave interference, which is caused by the superposition of echoes from different paths. Standing wave interference forms strip-like disturbances at equal depths on B-scan images, strongly suppressing valid data. In this paper, a mean image denoising algorithm is employed to reduce the impact of standing wave interference. First, calculate the sum of each column in each data block, and then determine the number of valid values. Subsequently, calculate the average value for each row, ignoring invalid or missing values, to obtain the average background noise for each row. Finally, subtract these averages from the original data to reduce the impact of background noise. Fig. 2(a) shows the comparison of the A-scan data before and after background denoising. By calculating the ratio of the power of the valid signal to the power of the noise signal, it was found that the signal-to-noise ratio of the data after denoising improved by approximately 189 %, demonstrating the effectiveness of this method

2. Zero offset removal. In the process of data collection, due to the extensive mileage of the railway lines and the prolonged detection time, the instability of the detection environment (such as changes in temperature and humidity) led to a zero offset of GPR data. Zero offset refers to the phenomenon where the A-scan data do not oscillate around the zero value and deviate from the theoretical baseline level. Zero offset can cause errors in estimating the depth of defect localization, reducing the accuracy of detection. To correct this deviation, this paper calculates the average value of the A-scan data and then subtracts this mean from the signal. Fig. 2(b) displays the comparison of the A-scan data before and after zero offset removal

*Data analysis*

After preprocessing the original data, a clear radar image is obtained. Fig. 3 displays the GPR grayscale image and A-scan data that reflect the structure of the ballast bed and subgrade. The 0–20 cm interval represents the air layer. The continuous jagged spike at 20 cm is caused by strong reflection from the rails and sleepers. From 20 to 50 cm, a series of irregular multiple reflections are observed, caused by reflections from ballast with different particle sizes. A clear in-phase axis at 50 cm marks the location of the subgrade surface layer.

For a ballast bed with uniform gradation and no obvious fouling, Eq. (2) indicates that the reflection coefficients of its internal regions are relatively consistent, and there are no distinct reflective layers in the GPR image, as shown in Fig. 4(a). Due to the higher relative permittivity of water (80) compared to that of ballast (4–8), the reflection coefficient $r_{j,j+1}$ increases on the surface of the wet ballast layer, resulting in a distinct reflection phenomenon. The radar image features of moisture are that the medium interface has low-frequency and strong reflection, large amplitude, contrary phase, and multiple reflections, as shown in Fig. 4(b). When the mud pumping occurs, a large amount of water-laden

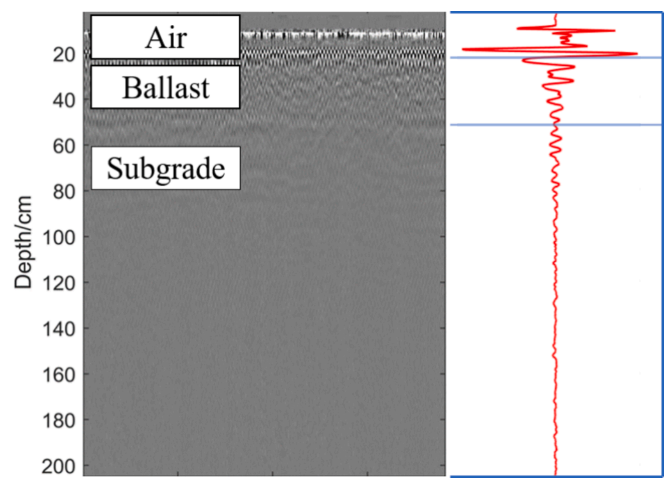
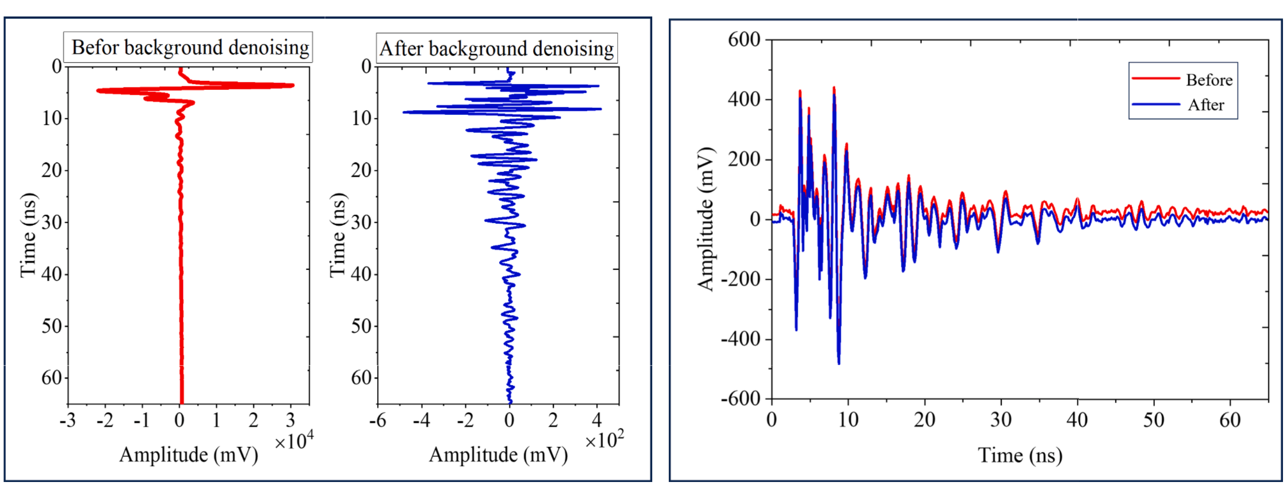
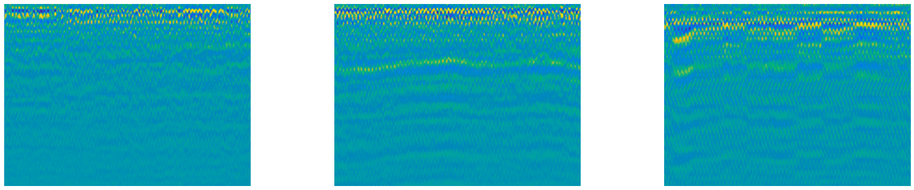



**Fig. 3.** GPR grayscale image profile.




(a) background denoising (b) zero offset removal

**Fig. 2.** Comparison of A-scan data before and after preprocessing.



(a) normal (b) moisture (c) mud pumping

**Fig. 4.** GPR profiles of railway ballast bed under different conditions.

silt rises to the surface of the ballast bed. Due to the varying water content in the mud, the radar image features of mud pumping are that the wave group of defects is disorderly and discontinuous, and the strong low-frequency reflection resembles a mountain tip or straw hat, as shown in Fig. 4(c).

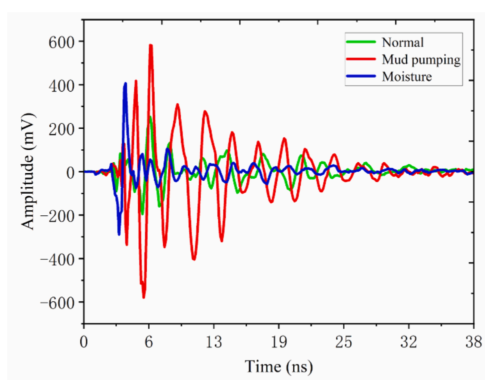
Given the non-stationary nature of GPR signals, which fluctuate with variations in the medium and over time, the STFT (Short-Time Fourier Transform) is employed to analyze these dynamic changes effectively. The principle of STFT involves segmenting the signal into overlapping short sections using a sliding window function and performing a discrete Fourier transform on each section. This process balances time and frequency resolution, allowing us to observe how the frequency components of a signal vary with time. The expression for the STFT is defined as[32]:

$$STFT_v(\tau,f) = \int_{-\infty}^{+\infty} x(t)g(t-\tau)e^{-j2\pi ft}d\tau \quad (4)$$

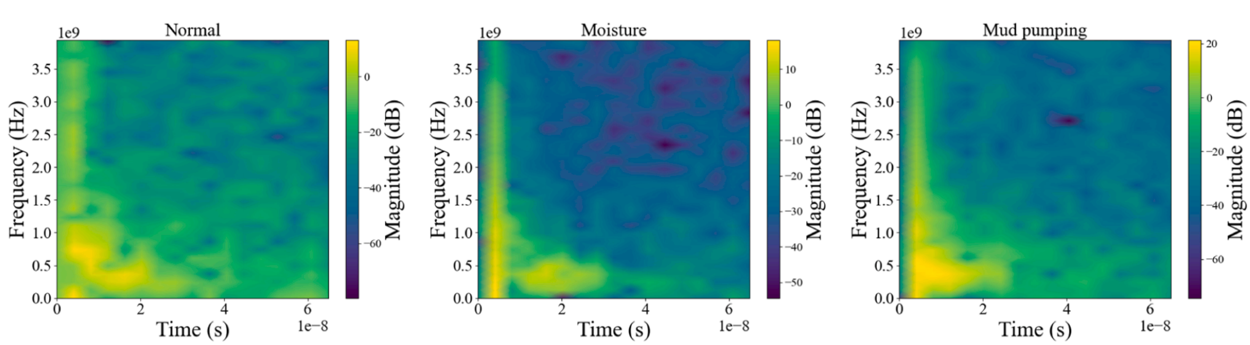
In the equation, $x(t)$ represents the time-domain signal, $g(t\text{-}\tau)$ is the window function used to select a specific time segment, and $\tau$ indicates the center position of the window. $f$ is the frequency, and $e^{-j2\pi ft}$ is the complex exponential form of the Fourier transform. Commonly used window functions include the rectangular window, triangular window, and Hamming window. Compared to other window functions, the Hamming window provides smoother boundaries, which can reduce spectral leakage, lower sidelobe levels, and improve signal frequency resolution. Therefore, this paper chooses the Hamming window function for the transformation. Fig. 5 shows the time–frequency differences among the three signals.

1.**Time domains.** As shown in Fig. 5(a), due to differences in water and fouling levels, there are significant variations among the three signals at the surface of the ballast bed (at 6 ns). The signal for mud pumping shows the peak at about 600 mV, the signal for moisture has a peak of about 400 mV, and the signal for normal ballast sections has the peak at approximately 250 mV, with their ratios roughly at 12:8:5.

2. **Frequency domains.** As shown in Fig. 5(b), because the central frequency of the GPR sampling is 400 MHz, the energy in the STFT image predominantly concentrates around this frequency. Due to the continuous attenuation of electromagnetic waves during propagation through the ballast and subgrade, the energy of the signal is primarily distributed between 0 and 20 ns. It can be deduced from Eq. (3) that electromagnetic waves attenuate more quickly in water. Therefore, the spectrogram for the normal shows the highest energy, whereas the




(a) comparison of three kinds of A-scan data in the time domain.




(b) comparison of three kinds of A-scan data in the frequency domain.

**Fig. 5.** Comparison of the differences between three kinds of A-scan data in both the time and frequency domains.

moisture shows the lowest energy.

## Methodology

### *Development of TFF-ResNet model*

The MLP (Multilayer Perceptron) is a classical type of multilayer feedforward neural network widely used in machine learning and deep learning fields, particularly suitable for performing classification tasks. The MLP typically includes an input layer, one or more hidden layers, and an output layer. For a three-class classification task, the output layer consists of three neurons. Each neuron outputs a value that, after being processed through a softmax function, represents the probability of a respective class, as shown in Fig. 6.

During model training, data is first entered into the input layer, then propagated forward through multiple hidden layers until it finally reaches the output layer. The output at each layer is derived by applying a weighted sum to the output from the preceding layer, adding a bias term, and subsequently processing this sum through an activation function, which can be represented as:

$$x^{(l)} = \sigma\left(W^{(l)}x^{(l-1)} + b^{(l)}\right) \tag{5}$$

where $x^{(l-1)}$ is the weighted output of the layer *l-1*, $W^{(l)}$ and $b^{(l)}$ are the weights and biases of the layer *l* respectively, σ is the activation function (such as Sigmoid or ReLU), and $x^{(l)}$ is the output of the layer *l*.

In this study, the cross-entropy loss (CE loss) function was employed to compute the discrepancy between predicted probabilities and actual class labels. The formulation of the loss function is given by the following expression:

$$L = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) \tag{6}$$

Where $C$ is the total number of classes, $C = 3$. $y_i$ is the one-hot encoded true labels, where if the sample belongs to the class $i$, then $y_i = 1$, otherwise $y_i = 0$. $\hat{y}_i$ is the predicted probability that the sample belongs to the class $i$, provided by the output of the softmax function.

ResNet18 significantly enhances the training efficiency and performance of deep networks through its innovative residual connections, excelling in image classification tasks. It begins with a 7x7 convolutional layer that processes the input image, followed by a batch normalization layer and a ReLU activation, and then a max pooling layer to reduce its spatial dimensions. Subsequently, the core of the network consists of residual blocks, which form the primary structure of the network. Each block contains two 3x3 convolutional layers. In each residual block, the input is not only processed through the convolutional layers, but also added directly to the output of the block via a shortcut connection. This design helps the network learn the residual function and assists in reducing the problem of vanishing gradients. The end of the network features a global average pooling layer, which compresses the output of the last residual block into a 1x1 plane. This is followed by a fully connected layer, with the number of nodes in the output layer equal to the number of classes in the classification task. The network structure of ResNet18 is shown in Fig. 7.

By combining the MLP and ResNet18 models, features are extracted from the A-scan data in both time and frequency domains. This bimodal structure, referred to as the TFF-ResNet (Time-Frequency-Fusion ResNet model) and illustrated in Fig. 8, integrates a sequence input layer for receiving A-scan data and an image input layer for spectrograms. The outputs from the MLP and ResNet18 layers are then merged into a longer vector by a feature concatenation layer. A fully connected layer then processes this merged vector to facilitate further analysis and classification.

GPR data are collected on multiple railway lines, and A-scan data for three types of ballast conditions—normal, moisture, and mud pumping—are extracted from the preprocessed data. The intelligent classification of ballast conditions is achieved using MLP, ResNet18, and the TFF-ResNet model developed in this paper, with the specific steps as follows:

1. **Data preparation.** Preprocess the original GPR data, which includes background denoising and zero offset removal, to enhance data quality. Extract A-scan data from the preprocessed data corresponding to normal, moisture, and mud pumping sections to create the dataset. The dataset contains a total of 55,920 A-scan signals, each consisting of 512 values. Divide the dataset from lines A, B, and C into training and validation datasets at a 7:3 ratio, with the data from line D serving as the testing dataset
2. **Model training.** The most suitable hyperparameter combinations for the MLP (number of hidden layers, number of neurons per layer, and learning rate) are determined using the random search algorithm [36]. During the time–frequency transformation of the A-scan data, the parameters of the STFT are continuously adjusted to modify the size and resolution of the spectrum, thereby enhancing the classification performance of the subsequent ResNet18. The best parameter combinations of these two models are then used to construct the TFF-ResNet model
3. **Model application.** The generalization ability of the model is verified using new data collected from another line, and the results of TFF-ResNet are compared with other machine learning-based models to validate the superiority of the method presented in this paper.

This experiment uses the deep learning framework PyTorch 1.10.0 + cu102 and Python 3.9.16 for programming. It was run on a desktop computer equipped with an Intel (R) Core (TM) i7-9700 CPU @ 3.00

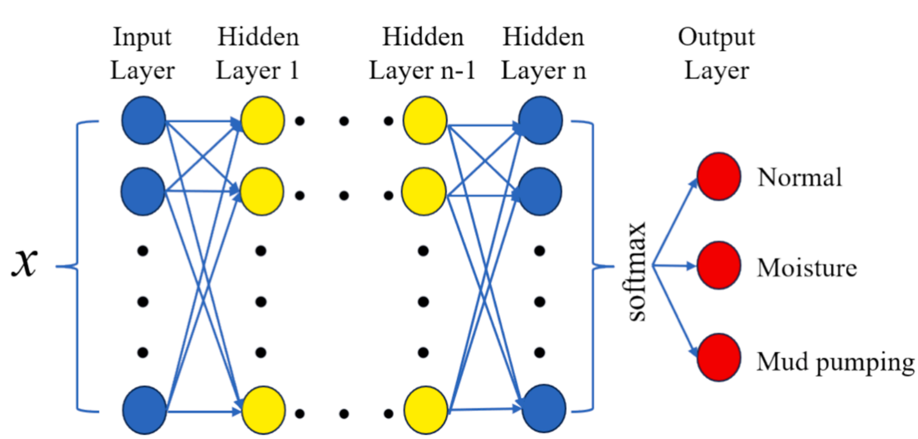



**Fig. 6.** Structure of MLP.

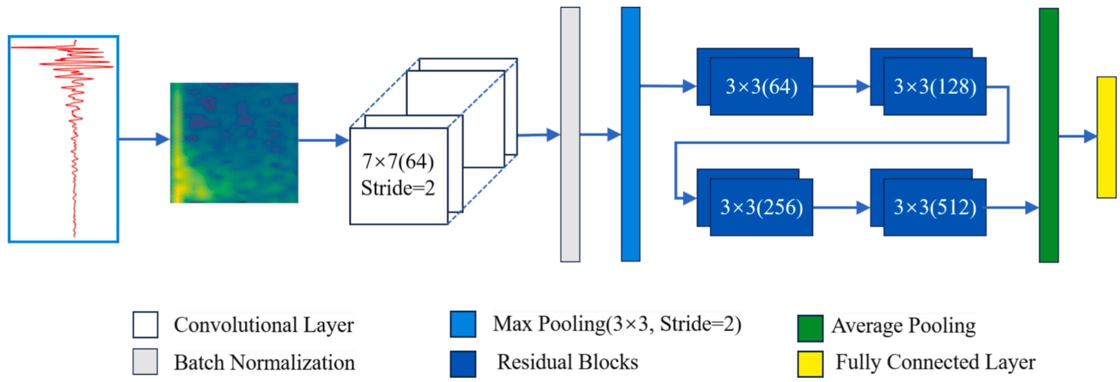



**Fig. 7.** Structure of ResNet18 model [35].

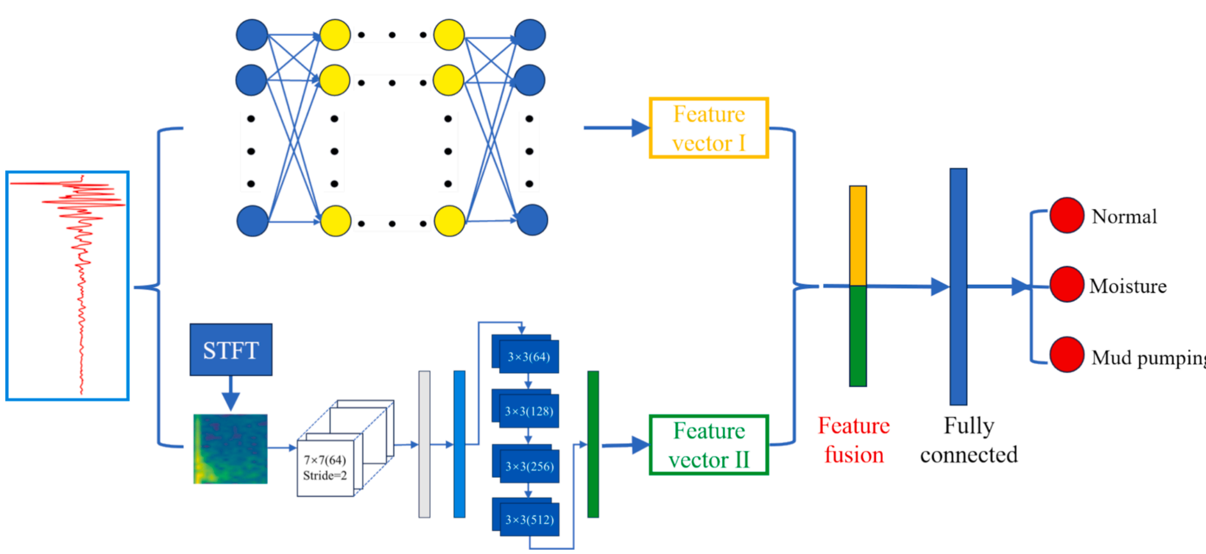



**Fig. 8.** Structure of TFF-ResNet model.

GHz, 24.0 GB of RAM, and an NVIDIA GeForce GTX 1660 SUPER graphics card. The method flowchart is shown in Fig. 9.

*Result*

For the MLP, an optimal combination of hyperparameters was determined using a random search algorithm. After 15 rounds of searching, the architecture was finalized with three hidden layers, each containing 820, 1060, and 960 neurons, respectively. In this paper, the A-scan data consists of 512 points. When performing the STFT, choosing factors of 512 as the window length avoids the need for data padding, thereby ensuring the accuracy of the analysis. Setting the overlap between windows to half the length of the time window provides a good balance between time and frequency resolution, reduces boundary effects, and makes the data analysis more stable and intuitive. Therefore, this study sets the range of window lengths to 16, 32, 64, 128, and 256, with corresponding overlaps between windows at 8, 16, 32, 64, and 128. The algorithm sequentially selects values to perform the STFT and inputs

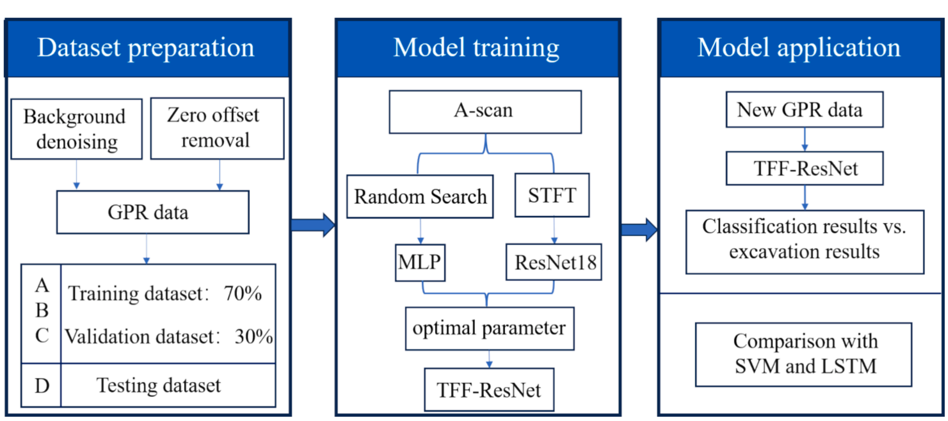



**Fig. 9.** Methodology flowchart.

the spectrograms into the ResNet18 model for training. By comparing the results from five groups, the optimal window length and overlap length are determined to be 64 and 32, respectively. The resultant spectrogram is a single-channel image of size 33x15, which is then replicated into a 3-channel image before being fed into the ResNet18 network. Based on the above model parameters, the structure of TFF-ResNet is constructed. The training parameters and other structural details of each model are outlined in Table 3.

In Fig. 10, it can be observed that the CE losses of the three models gradually approach the minimum value and converge after approximately 200, 100, and 200 epochs, respectively. Therefore, 500 epochs are sufficient to train the models and obtain reliable results. Furthermore, each model performs well on the validation dataset, closely matching the performance on the training dataset, indicating that the models can avoid the problem of overfitting.

Fig. 11 shows the classification performance of the three models using ROC curves. Overall, the three ROC curves of the TFF-ResNet are closest to the top left corner and therefore perform best. Fig. 11(a) reveals that the ResNet18 model outperforms the MLP on the validation dataset. Conversely, Fig. 11(b) indicates that the MLP slightly surpasses the ResNet18 model on the testing dataset.

In order to quantitatively evaluate the classification performance of the three models, confusion matrixes are used for representation, as shown in Fig. 12. The numbers 0, 1, and 2 are used to represent normal, moisture, and mud pumping, respectively. From these confusion matrices, we can derive the accuracy (A), macro-average precision (P), macro-average recall (R), and macro-average F1 score (F). Accuracy is defined as the ratio of the number of correctly predicted samples to the total number of samples, as shown in Eq. (7); macro-average precision calculates the precision for each category individually and then averages these values, as shown in Eq. (8); macro-average recall calculates the recall for each category individually and then averages these values, as shown in Eq. (9); macro-average F1 score calculates the F1 score for each category individually and then averages these values, as shown in Eq. (10).

$$A = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

$$P = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i + FP_i} \tag{8}$$

$$R = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i + FN_i} \tag{9}$$

$$F = \frac{1}{N}\sum_{i=1}^{N}2\times\frac{P_i \times R_i}{P_i + R_i} \tag{10}$$

Where $TP$, $TN$, $FP$ and $FN$ represent the true positives, true negatives, false positives, and false negatives for the entire dataset, respectively. $TP_i$, $TN_i$, $FP_i$ and $FN_i$ represent the true positives, true negatives, false positives, and false negatives for the *i-th* class of samples ($i = 1, 2, 3$). $P_i$ and $R_i$ are the precision and recall for the *i-th* class of samples. The specific values are listed in Table 4.

The TFF-ResNet model demonstrated significant performance improvements over the MLP and ResNet18 models in both the validation and testing datasets. In the validation dataset, the TFF-ResNet model achieved the highest metrics with an accuracy of 0.878, macro-average precision of 0.875, macro-average recall of 0.875, and macro-average F1 score of 0.875. It showed an average improvement of approximately 24 % over the MLP and 14 % over the ResNet18 model. In the testing dataset, compared to the MLP, the TFF-ResNet model's performance improved by about 21 %, and relative to the ResNet18 model, it exhibited a 33 % improvement in accuracy, 27 % in precision, 41 % in recall and 34 % in F1 score. Furthermore, all models performed better on the validation dataset than on the testing dataset, posing a challenge to the models' generalization capabilities.

In order to observe the classification performance of the three models on each type of ballast bed defect in more detail, the F1 score is used as the evaluation index, and the comparison effect is shown in Fig. 13.

In the validation dataset, the TFF-ResNet model consistently achieved higher F1 scores across all ballast bed conditions compared to the other two models. Notably, in recognizing the normal state, the TFF-ResNet model reached an F1 score of 0.9, which is significantly better than the F1 scores of the MLP at 0.76 and the ResNet18 model at 0.83. In the testing dataset, all models experienced a performance decline, likely due to variations in climate conditions, soil types, construction materials, and building standards across different regions, resulting in discrepancies between the samples in the validation and testing datasets. Despite these variations, the TFF-ResNet model still maintained the highest F1 score in the testing dataset.

*Validation and comparison*

To further validate the wide applicability of the proposed model, this paper selected a general-speed railway in Southern China as the experimental interval, with data acquisition parameters consistent with those in Table 1. To prevent the model from mistakenly classifying other ballast bed conditions into these three categories, a decision threshold of 0.6 was set. That is, if the maximum prediction probability of the model for a sample is less than 0.6, the model will judge this sample as the fourth type and output label 3. This fourth type of label corresponds to bridges, turnouts, tunnels and other structures on the railway line, or other ballast bed defects, such as ballast bed subsidence, frost heaving and so on. Fig. 14 (a) shows the collection of GPR grayscale images of three ballast bed defects and bridges, while Fig. 14 (b) shows the corresponding classification result. The result indicates that the model can effectively differentiate between the four ballast bed conditions. The overall classification accuracy is 0.75, which is lower than that of the testing dataset, possibly due to the setting of the decision threshold and the introduction of the fourth type of sample. The data used in the above verification includes a total of 400 samples, and the output result of the model takes only 0.8 s, which ensures the real-time performance of the model detection.

In order to verify the superiority of the TFF-ResNet model, it was compared with other machine learning-based classification models. The SVM algorithm excels at handling data within high-dimensional spaces and can effectively manage complex nonlinear feature relationships using different kernel functions, making it widely used in classifying A-scan signals from GPR. LSTM, with its unique gating mechanism and capabilities for long-term dependency handling, is extensively used for processing and predicting time-series data. A-scan data inherently constitutes time-series data, as it records the propagation time and signal strength of radar waves at various depths. Consequently, this chapter chose SVM and LSTM for performance comparison with the TFF-ResNet, with the results presented in Table 5.

The table shows that the classification performance of the three models on the testing dataset is lower than on the validation dataset, consistent with the conclusions in Section 3.2. On the validation dataset, all SVM metric values range between 0.65 and 0.69, approximately 23.2 % lower than those of the TFF-ResNet model. The LSTM model shows

**Table 3**
Partial structural and training parameters of the three models.

| Parameters | MLP | ResNet18 | TFF-ResNet |
|---|---|---|---|
| Learning rate | 0.001 | 0.0001 | 0.0001 |
| Batch size | 32 | 16 | 16 |
| Epochs | 500 | | |
| Loss function | CE loss | | |
| Activation function | ReLu | | |
| Optimiser | Adam | | |

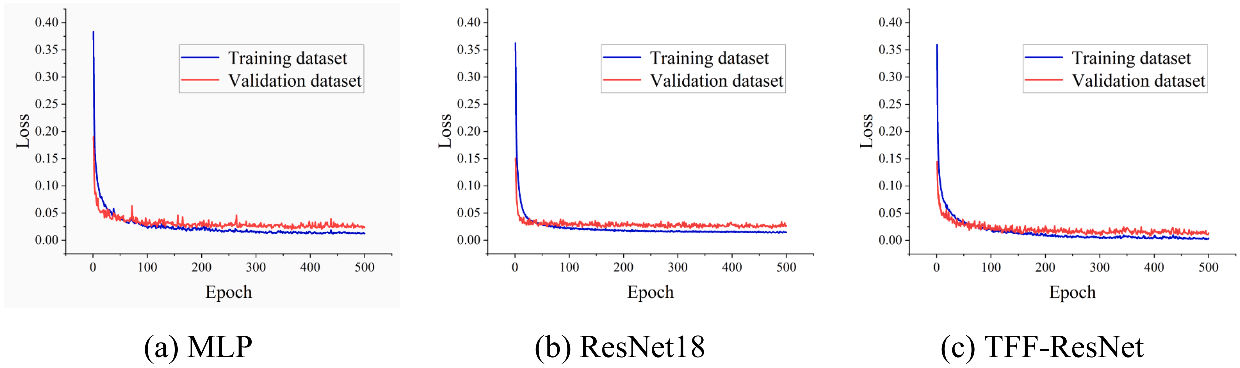



**Fig. 10.** Comparison of CE losses of training and validation datasets during training process.

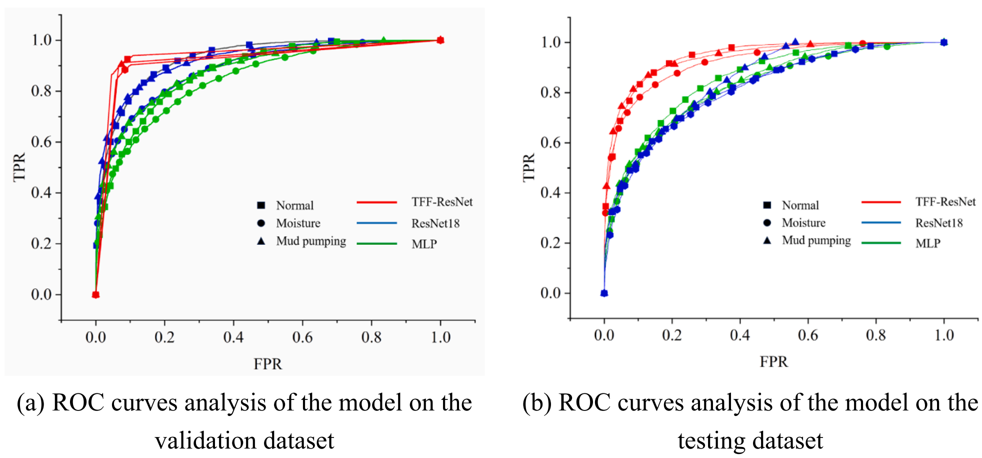



**Fig. 11.** ROC curves obtained by three models on the dataset.

the lowest performance, all metric values range between 0.60 and 0.65, approximately 28.5 % lower than those of the TFF-ResNet model. On the testing dataset, all SVM metric values range between 0.62 and 0.67, while LSTM has a very poor classification effect.

## Conclusions

Due to the acceleration of globalization and the continuous growth of international trade, the strategic position and importance of rail transport have significantly increased. As rail transport volume continues to grow, issues with ballast bed defects affecting railway safety have become increasingly prominent, with moisture and mud pumping in the ballast bed being especially common. Vehicle-mounted GPR is a frequently used method for detecting these ballast bed defects. To avoid the low automation associated with traditional manual interpretation of radar images to determine the type and location of defects, many machine learning methods have been developed for the intelligent classification and recognition of GPR data. However, these methods often involve cumbersome initial work or suffer from poor generalization. This paper introduces a bimodal deep learning classification algorithm with two channels: one uses a MLP to extract features from time-domain signals, and the other applies STFT and a ResNet18 to extract features from frequency-domain signals. Results indicate that the bimodal model, TFF-ResNet, outperforms standalone MLP and ResNet18 models in classification performance on both validation and testing datasets.

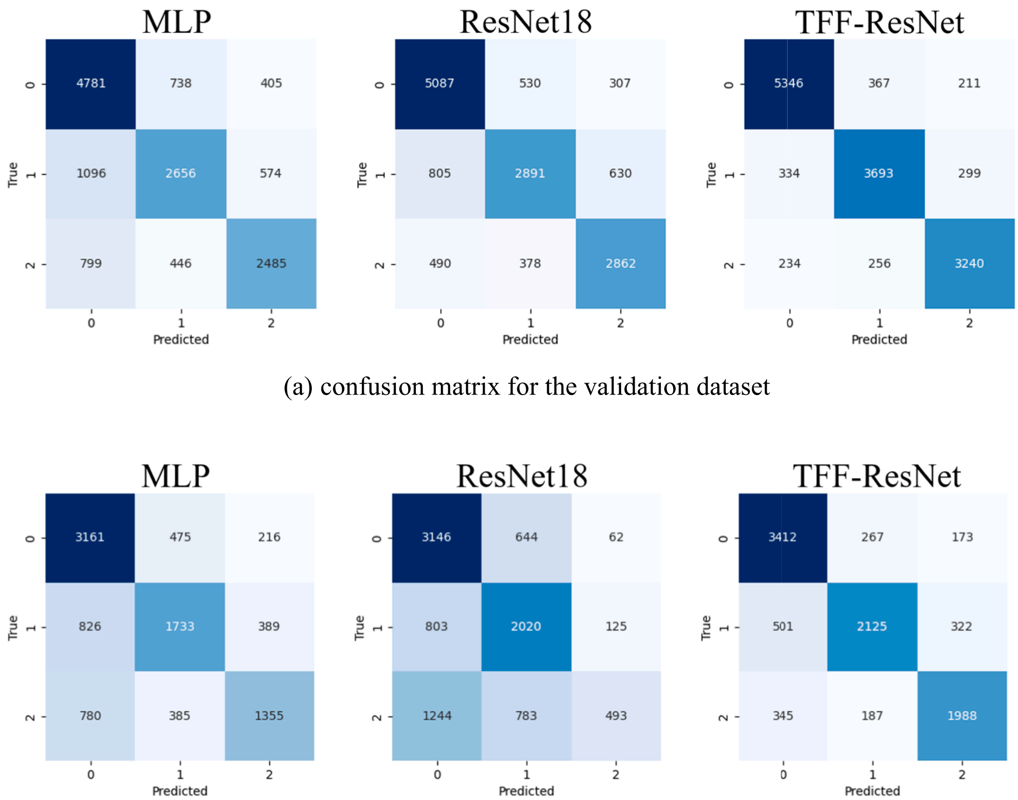
After multiple rounds of experiments, this paper mainly draws the following conclusions:

1. The standalone MLP exhibits lower classification performance on the validation dataset compared to the ResNet18; however, it demonstrates superior generalization on the testing dataset over the ResNet18
2. By extracting features from both the time domain and frequency domain, the TFF-ResNet model achieved approximately 24 % and 14 % improvements in classification performance on the validation dataset compared to the MLP and ResNet18 models, respectively
3. The TFF-ResNet model demonstrates superior generalization capabilities on other railway lines, achieving a 21 % and 34 % improvement in classification performance on the testing dataset relative to the MLP and ResNet18 models, respectively

This model is primarily developed for detecting moisture and mud pumping defects in railway ballast beds. Currently, it has not been developed further to address other types of defects, such as subsidence and frost heaving, due to insufficient data available for training the model on these issues. In our future research, we aim to expand the model's detection capabilities by incorporating more diverse defect data. Additionally, future efforts will focus on exploring the combination of this method with other detection technologies, such as thermal imaging and electrical resistivity tomography, to create a multimodal integrated defect detection system that improves the comprehensiveness and accuracy of inspections. The model has the potential to be integrated into intelligent railway infrastructure maintenance systems, providing real-time ballast bed condition monitoring and defect alerts, therefore, enhancing the efficiency and safety of railway maintenance.

## CRediT authorship contribution statement

**Junjie Bu:** Writing – original draft, Visualization, Methodology, Formal analysis, Conceptualization. **Guoqing Jing:** Writing – review &




(a) confusion matrix for the validation dataset

(b) confusion matrix for the testing dataset

**Fig. 12.** Confusion matrices obtained by three models on the dataset.

**Table 4**
The overall classification performance of each model.

| Model | validation dataset | | | | testing dataset | | | |
|---|---|---|---|---|---|---|---|---|
| | A | P | R | F | A | P | R | F |
| MLP | 0.710 | 0.708 | 0.696 | 0.702 | 0.670 | 0.674 | 0.649 | 0.661 |
| ResNet18 | 0.775 | 0.770 | 0.765 | 0.768 | 0.607 | 0.639 | 0.566 | 0.600 |
| TFF-ResNet | 0.878 | 0.875 | 0.875 | 0.875 | 0.807 | 0.809 | 0.798 | 0.804 |

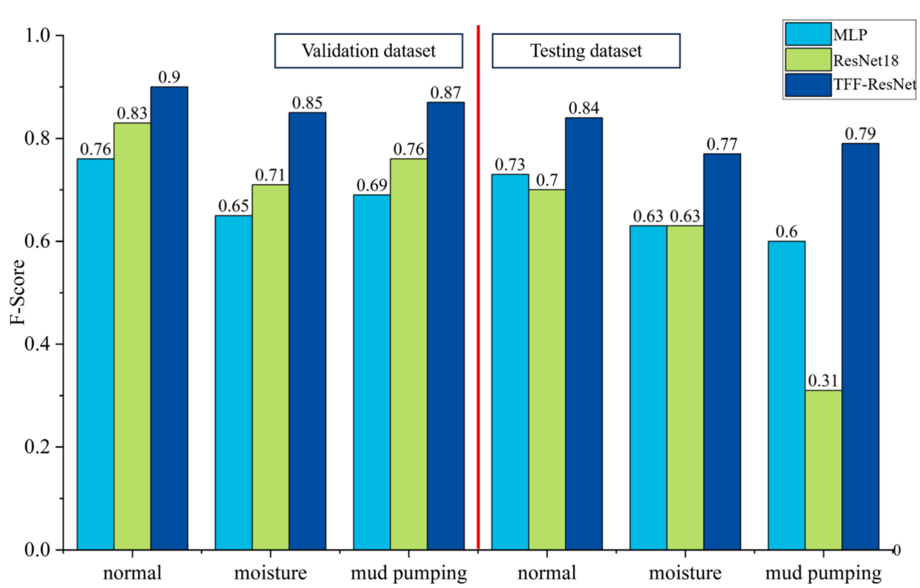
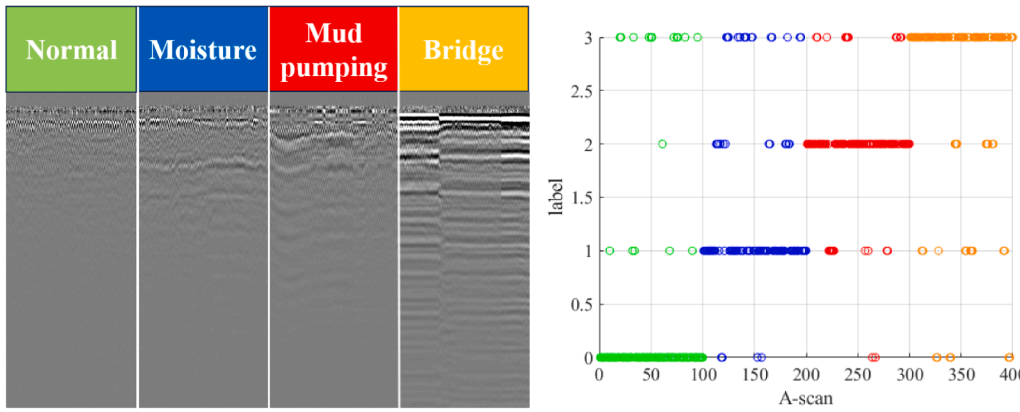



**Fig. 13.** Comparison of classification performance for three kinds of defects using F-Score as the evaluation metric.




(a) sample for verification (b) verification results

**Fig. 14.** Using GPR data from the fifth railway line to validate the generalization ability of model.

**Table 5**
Comparison of classification performance of three machine learning-based models.

| Model | validation dataset | | | | testing dataset | | | |
|---|---|---|---|---|---|---|---|---|
| | A | P | R | F | A | P | R | F |
| TFF-ResNet | 0.878 | 0.875 | 0.875 | 0.875 | 0.807 | 0.809 | 0.798 | 0.804 |
| SVM | 0.685 | 0.683 | 0.658 | 0.663 | 0.661 | 0.663 | 0.622 | 0.625 |
| LSTM | 0.636 | 0.645 | 0.609 | 0.613 | 0.567 | 0.491 | 0.501 | 0.445 |

editing, Funding acquisition. **Xujie Long:** Software. **Lei Wang:** Supervision. **Zhan Peng:** Resources, Project administration, Investigation, Data curation. **Yunlong Guo:** Validation.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgement

This work was financially supported by the National Natural Science Foundation of China (No. 52027813); Research on the Overhaul of Heavy-Duty Railway Technical System and Key Technologies (No. 2023YJ193); Rapid Collaborative Analysis Technology for Ballast Track Subgrade Georadar Data (No. 2022YJ305). The European Commission and UKRI Engineering and Physical Science Research Council (EPSRC) are acknowledged for the financial sponsorship of Re4Rail project (Grant No EP/Y015401/1).

## Data availability

The data that has been used is confidential.

## References

[1] Jing GQ. Railway Ballasted Bed. Beijing: China Railway Publishing House; 2012.
[2] Wang SL, Liu GX, Jing GQ, et al. State-of-the-art review of ground penetrating radar (GPR) applications for railway ballast inspection. Sensors 2022;22(7):2450. https://doi.org/10.3390/s22072450.
[3] Wan ZB, Bian XC, Chen YM. Mud pumping in high-speed railway: in-situ soil core test and full-scale model testing[J]. Railway Eng Sci 2022;30(3):289–303. https://doi.org/10.1007/s40534-022-00283-0.
[4] Wang W, Wu YJ, Yang CZ, Feng QS. Mechanism and influencing factors of mud pumping in railway subgrade. J Traffic Transp Eng 2019;19(6):54–64. https://doi.org/10.19818/j.cnki.1671-1637.2019.06.006.
[5] Nie RS, Li WM, Su Y, et al. Physical and mechanical properties of mud pumping soils in railway subgrade bed. J Southwest Jiaotong Univ 2018;53(2):286–95. https://doi.org/10.3969/j.issn.0258-2724.2018.02.010.
[6] Huang J, Su Q, Wang W, et al. Field investigation and full-scale model testing of mud pumping and its effect on the dynamic properties of the slab track–subgrade interface. Proc Inst Mech Eng Part F J Rail Rapid Transit 2019;233(8):802–16. https://doi.org/10.1177/0954409718810262.
[7] Huang J, Su Q, Wang W, et al. Vibration behavior and reinforcement effect analysis of the slab track-subgrade with mud pumping under cyclic dynamic loading: full-scale model tests. Shock Vib 2018;2018:3087254. https://doi.org/10.1155/2018/3087254.
[8] Wang SL, Zhao ZG, Liu HJ, et al. Research and application progress of radar detection technology for railway ballast bed. J Railway Soc 2023;45(2):116–26. https://doi.org/10.3969/j.issn.1001-8360.2023.02.013.
[9] Zhu DB, Gao D, Lai QL, et al. Study on influence of main frequency of vehicle-borne ground penetrating radar on evaluation parameters of ballast bed contamination. J China Railway Soc 2022;44(12):97–104. https://doi.org/10.3969/j.issn.1001-8360.2022.12.012.
[10] Silvast M, Levomäki M, Nurmikolu A, et al. NDT techniques in railway structure analysis. In: Proceedings of the 7th World Congress on Railway Research, Montreal, Canada, June 4-8, 2006. 2006; p. 12.
[11] Leng Z, Al-Qadi IL. Railroad ballast evaluation using ground-penetrating radar: Laboratory investigation and field validation. Transp Res Rec 2010;2159(1):110–7. https://doi.org/10.3141/2159-14.
[12] Al-Qadi IL, Zhao S, Shangguan P. Railway ballast fouling detection using GPR data: Introducing a combined time–frequency and discrete wavelet techniques. Near Surf Geophys 2016;14(2):145–53. https://doi.org/10.3997/1873-0604.2016006.
[13] Wang SL, Peng Z, Li M, et al. Research on state characterization index of railway ballast bed based on high frequency radar. China Railw Sci 2022;43(3):8–16. https://doi.org/10.3969/j.issn.1001-4632.2022.03.02.
[14] Benedetto F, Tosti F, Alani AM. An entropy-based analysis of GPR data for the assessment of railway ballast conditions. IEEE Trans Geosci Remote Sens 2017;55(7):3900–8. https://doi.org/10.1109/TGRS.2017.2683507.
[15] Benedetto A, Tosti F, Ciampoli LB, et al. An overview of ground-penetrating radar signal processing techniques for road inspections. Signal Process 2017;132:201–9. https://doi.org/10.1016/j.sigpro.2016.05.016.
[16] Tosti F, Adabi S, Pajewski L, et al. Large-scale analysis of dielectric and mechanical properties of pavement using GPR and LFWD. In: Proceedings of the 15th International Conference on Ground Penetrating Radar. IEEE; 2014:868-873. https://doi.org/10.1109/ICGPR.2014.6970551.
[17] Shihab S, Al-Nuaimy W, Huang Y, et al. Neural network target identifier based on statistical features of GPR signals. In: Proceedings of the Ninth International Conference on Ground Penetrating Radar. SPIE; 2002;4758:135-138. https://doi.org/10.1117/12.462228.
[18] Gao YT, Xu J, Wang YH, et al. Depth estimation of buried structures based on the GPR reflected waveform characteristics. Chin J Eng 2016;38(12):1798–805. https://doi.org/10.13374/j.issn2095-9389.2016.12.020.
[19] Liu CL, Du YC, Yue GH, et al. Advances in automatic identification of road subsurface distress using ground penetrating radar: State of the art and future trends. Autom Constr 2024;158:105185. https://doi.org/10.1016/j.autcon.2023.105185.

[20] Sezgin M, Yoldemir B, Özkan E, et al. Identification of buried objects based on peak scatter modelling of GPR A-scan signals. In: Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XXIV. SPIE; 2019;11012:39-48. https://doi.org/10.1117/12.2517691.
[21] Gou JP, Ma HX, Ou WH, et al. A generalized mean distance-based k-nearest neighbor classifier. Expert Syst Appl 2019;115:356–72. https://doi.org/10.1016/j.eswa.2018.08.021.
[22] Shao WB, Bouzerdoum A, Phung SL, et al. Automatic classification of ground-penetrating-radar signals for railway-ballast assessment. IEEE Trans Geosci Remote Sens 2011;49(10):3961–72. https://doi.org/10.1109/TGRS.2011.2128328.
[23] Du PF, Liao LJ, Yang XA. Intelligent recognition of defects in railway subgrade. J China Railway Soc 2010;32(3):142–6. https://doi.org/10.3969/j.issn.1001-8360.2010.03.025.
[24] Xu JC, Zhang JK, Sun WG. Recognition of the typical distress in concrete pavement based on GPR and 1D-CNN. Remote Sens 2021;13(12):2375. https://doi.org/10.3390/rs13122375.
[25] Ahmadvand M, Dorafshan S, Azari H, et al. 1D-CNNs for autonomous defect detection in bridge decks using ground penetrating radar. In: Health Monitoring of Structural and Biological Systems XV. SPIE; 2021;11593:97-113. https://doi.org/10.1117/12.2580575.
[26] Liu H, Wang SL, Jing GQ, et al. Combined CNN and RNN neural networks for GPR detection of railway subgrade diseases. Sensors 2023;23(12):5383. https://doi.org/10.3390/s23125383.
[27] Özkaya U, Öztürk Ş, Melgani F, et al. Residual CNN+ Bi-LSTM model to analyze GPR B scan images. Autom Constr 2021;123:103525. https://doi.org/10.1016/j.autcon.2020.103525.
[28] Xu XJ, Lei Y, Yang F. Railway subgrade defect automatic recognition method based on improved faster R-CNN. Sci Program 2018;2018:4832972. https://doi.org/10.1155/2018/4832972.
[29] Ma ZX, Yang F, Qiao X. Intelligent detection method of railway subgrade defect. Comput Eng Appl 2021;57(9):272–8. https://doi.org/10.3778/j.issn.1002-8331.2002-0076.
[30] Lei WT, Luo JB, Hou FF, et al. Underground cylindrical objects detection and diameter identification in GPR B-scans via the CNN-LSTM framework. Electronics 2020;9(11):1804. https://doi.org/10.3390/electronics9111804.
[31] Zhang J, Zhang C, Lu Y, et al. In-situ recognition of moisture damage in bridge deck asphalt pavement with time-frequency features of GPR signal[J]. Constr Build Mater 2020;244:118295.
[32] Al-Qadi IL, Xie W, Roberts R, et al. Data analysis techniques for GPR used for assessing railroad ballast in high radio-frequency environment[J]. J Transp Eng 2010;136(4):392–9.
[33] Wang H, Ouyang S, Liao KF, et al. GPR B-SCAN image hyperbola detection method based on deep learning. Acta Electron Sin 2021;49(5):953–63. https://doi.org/10.12263/DZXB.20200635.
[34] Wang H, Ouyang S, Liu QH, et al. Structure feature detection method for ground penetrating radar two-dimensional profile image based on deep learning. J Electron Inf Technol 2022;44(4):1284–94. https://doi.org/10.11999/JEIT211032.
[35] Chotikunnan P, Puttasakul T, Chotikunnan R, et al. Evaluation of single and dual image object detection through image segmentation using ResNet18 in robotic vision applications. J Robotics Control 2023;4(3):263–77. https://doi.org/10.18196/jrc.v4i3.17932.
[36] Han L, Liao Y, Wang H, et al. Long-term prediction for railway track geometry based on an optimised DNN method. Constr Build Mater 2023;401:132687. https://doi.org/10.1016/j.conbuildmat.2023.132687.