TU Delft

Delft University of Technology

Document Version
Final published version

Licence
Dutch Copyright Act (Article 25fa)

Citation (APA)
Yue, Y., Liu, H., Liu, X., Chen, Z., Li, Y., Lin, P., & Cui, J. (2026). Rapid Non-Contact Road Surface Inspection Using Tire-Road Coupling Noise and MFCC feature. *Advanced Engineering Informatics*, *74*, Article 104774. https://doi.org/10.1016/j.aei.2026.104774

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

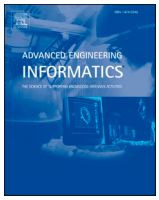
Advanced Engineering Informatics 74 (2026) 104774



Contents lists available at ScienceDirect

Advanced Engineering Informatics

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



Rapid Non-Contact Road Surface Inspection Using Tire-Road Coupling Noise and MFCC feature

Yunpeng Yue [a], Hai Liu [a,*], Xiaoyu Liu [a], Zhijie Chen [a], Yi Li [b], Peng Lin [c], Jie Cui [a,*]

[a] School of Civil Engineering and Transportation, Guangzhou University, Guangzhou 510006, China
[b] Section of Pavement Engineering, Faculty of Civil Engineering & Geosciences, Delft University of Technology, Stevinweg 1, 2628 CN Delft, The Netherlands
[c] Rijkswaterstaat, Ministry of Infrastructure and Water Management, 3526 LA Utrecht, The Netherlands

ARTICLE INFO

Keywords:
Road inspection
Defect
Deep learning
Mel-frequency cepstral coefficients
Tire-road coupling noise

ABSTRACT

Accurate and efficient inspection of road surface defects is essential for ensuring traffic safety and supporting timely maintenance. However, existing vision-based inspection techniques often rely on high-resolution cameras or laser sensors, which limit robustness under complex or dynamic environmental conditions. To overcome these challenges, this study develops a vehicle-mounted acoustic inspection framework based on tire–road coupling noise for rapid road surface condition assessment. An acoustic inspection dataset is established, covering multiple types of urban pavement defects under diverse traffic and environmental conditions. The captured noise signals are transformed into MFCC, which serve as frequency-domain features for defect detection. A deep neural network integrating CNN, SE attention, and BiLSTM modules is developed to extract multi-scale time-frequency features and model temporal dependencies from tire-road coupling noise for pavement defect detection. Model evaluation results demonstrate that the proposed CNN-SE-BiLSTM model can accurately classify pavement defects without reliance on visual sensors, achieving an overall F1-score of 84.0%, while maintaining real-time inference and strong robustness to varying road surface conditions. Compared with existing vision- and Sensor- based inspection methods, the proposed method offers advantages in lower deployment cost, reduced sensitivity to illumination and weather variations, and easier integration into standard vehicles for continuous large-scale inspection. A field experiment on urban roads verifies the effectiveness of the proposed road surface inspection method, and a total of 12 pavement defects, including 9 cracks and 3 potholes, were successfully identified under real driving conditions. It is concluded that the proposed tire-road coupling noise method provides a cost-effective solution for road surface inspection.

1. Introduction

Pavement inspection plays a critical role in ensuring road safety, maintaining structural integrity, and supporting timely maintenance decision-making for modern transportation networks [1,2]. With increasing urban traffic volume, temperature fluctuations, and water-induced erosion, pavement structures suffer from defects such as cracks, potholes, and material aging [3,4]. These defects not only degrade driving comfort but also pose threats to transportation safety and roadway mobility [5,6]. Although traditional manual inspection methods are still widely adopted, their inherent time consumption, labor intensity, and reliance on inspector expertise restrict efficient application to large-scale and rapidly evolving urban road networks [7,8]. Therefore, development of automated, reliable, and scalable pavement inspection technologies has become an urgent engineering necessity.

Traditional non-contact inspection technologies, such as computer vision (CV) [9], ground penetrating radar (GPR) [10], and 3D laser scanning [11], have been widely adopted to improve detection efficiency and reduce reliance on manual labor. Recent studies on pavement condition inspection have explored a range of sensing techniques, each offering distinct advantages arising from their underlying physical mechanisms. As summarized in Table 1, image-based techniques dominate the existing literature, benefiting from high spatial resolution and the widespread adoption of Convolutional neural networks (CNNs). Image-based techniques employing CCD cameras, stereo vision, or UAV platforms generally report high detection accuracy, particularly for surface-level defects such as cracks and potholes. Sensor-based techniques, including GPR, infrared thermography, and LiDAR, extend

* Corresponding author.
E-mail addresses: hliu@gzhu.edu.cn (H. Liu), jcui2009@hotmail.com (J. Cui).

https://doi.org/10.1016/j.aei.2026.104774
Received 8 March 2026; Received in revised form 13 April 2026; Accepted 29 April 2026
Available online 8 May 2026
1474-0346/© 2026 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Table 1
Comparison of deep learning models for road surface detection using various techniques.

Type	Reference	Year	Device	Algorithm	Detection Type	Inspection result
Image-based Detection	Zhang et al [44]	2023	Digital camera	YOLOX	Pothole	98.1% F1-score
	Zhai et al [45]	2023	Digital camera	Improved Faster R-CNN	Longitudinal crack, Transverse crack, Alligator crack	87.8% mAP
	Liu et al [46]	2024	CD455 visual sensor	MASL-YOLO	Lateral crack, Longitudinal crack, Pothole, Alligator crack, Manhole distress	84.4% Precision
	Jiang et al [7]	2024	UAV-based visual sensor	YOLOv7-RDD	Pothole	69.0% F1-score
Sensor-based Detection	Liu et al [47]	2023	GPR	YOLOv3-FDL	Crack	86.3% Precision
	Xiong et al [12]	2024	GPR	YOLOv3	Crack, Debonding damage	96.9% Accuracy
	Jiang et al [48]	2022	Infrared camera	GSkYOLOv5	Cross,Reticular, Strip	77.2% Precision
	Karukayil et al [49]	2024	LIDAR	YOLOv5	Pothole	86.0% Precision

inspection capabilities beyond surface appearance by capturing subsurface anomalies, thermal responses, and layer discontinuities. Owing to their ability to probe internal pavement structures, these techniques can achieve higher detection accuracy, in some cases approaching 96.9% [12]. However, these performances typically come at the cost of increased system complexity, specialized hardware, and higher deployment requirements.

Tire-road coupling signals have emerged as an alternative information source for pavement condition inspection [13,14]. Tire-road coupling noise is generated by the dynamic coupling between vehicle tires and pavement macro-texture, rendering it inherently sensitive to surface defects such as crack and pothole [15,16]. Compared with vision- or sensor-based techniques, acoustic sensing exhibits high robustness to variations in dust, rainfall and lighting conditions [17]. Moreover, acoustic-based techniques are low-cost, lightweight, and can be readily integrated into ordinary vehicles, enabling continuous data acquisition during normal driving without the need for specialized equipment [18,19]. This capability enables tire-road noise signals to be collected at relatively high speeds over extended distances, offering potential support for rapid and non-contact pavement inspection in urban road networks [20].

Although tire–road coupling noise contains rich acoustic information, extracting stable defect-related patterns is challenging due to its dynamic, nonlinear nature and sensitivity to operating conditions [21,22]. This challenge motivates the use of dedicated acoustic feature learning models to capture temporal-spectral characteristics of tire-road coupling noise [23]. Existing acoustic-based studies often rely on raw waveform features or alternative representations such as spectrograms or wavelet-based features. For example, Raslan et al. [24] employed time–frequency representations such as spectrograms for road condition analysis, while Kamranfar et al. [25] have explored wavelet-based or perceptual linear prediction features. However, these representations may not effectively capture perceptually relevant spectral characteristics of tire–road interaction signals. In contrast, MFCC features provide a compact and discriminative representation of acoustic signals by modeling frequency responses on the mel scale, which better aligns with human auditory perception and enhances the representation of subtle acoustic variations induced by pavement defects [26,27]. In recent years, the development of deep learning methods has further advanced acoustic feature modeling in areas such as audio classification, mechanical fault diagnosis, and structural health monitoring [28]. CNNs and recurrent neural networks (RNNs), including long short-term memory (LSTM) and gated recurrent unit, have demonstrated strong capabilities in extracting discriminative patterns from time–frequency representations [29,30]. Recent studies have also explored lightweight CNN architectures to improve computational efficiency and deployment flexibility in various engineering applications, highlighting their potential for real-time and resource-constrained scenarios [31–33]. By integrating MFCC features with deep learning models, more robust temporal–spectral patterns associated with pavement defects can be effectively captured [34,35]. In addition, attention mechanisms can enhance acoustic feature learning by emphasizing informative signal components [36,37]. Despite these advances, deep learning applications to tire–road coupling noise remain limited, particularly across different pavement types.

Existing studies on pavement condition inspection primarily rely on

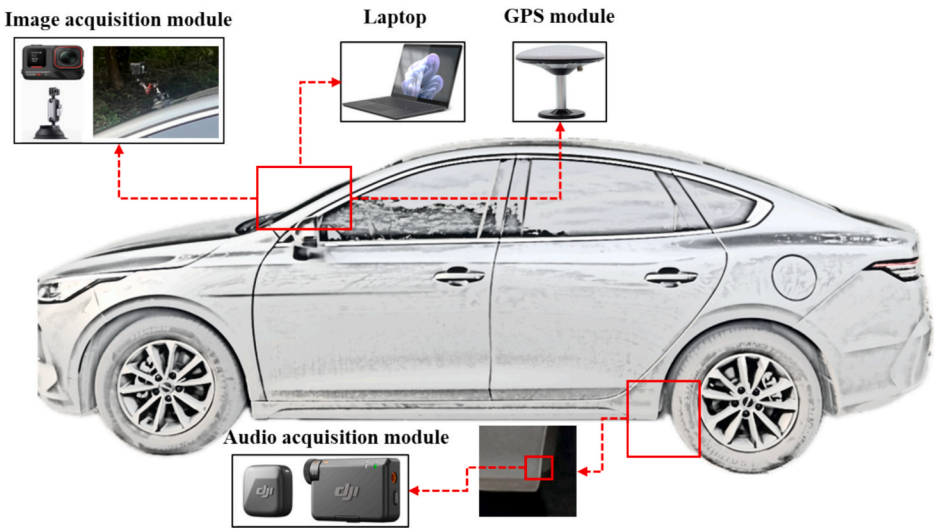


Fig. 1. Vehicle-mounted data acquisition setup for tire-road coupling noise.

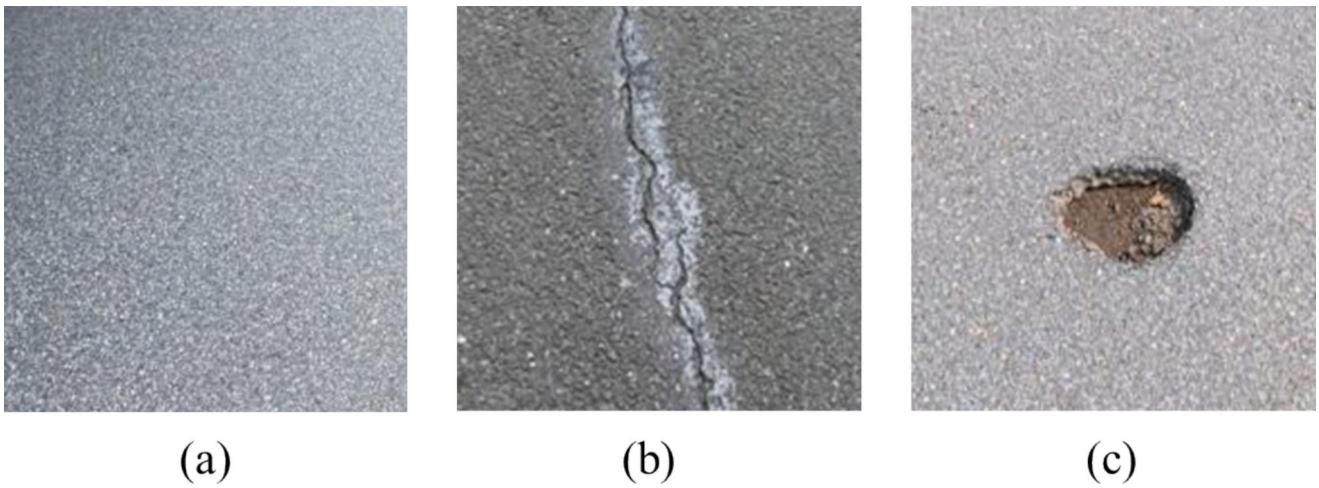


Fig. 2. Representative examples of different types of road surface defects: (a) normal, (b) crack, and (c) pothole.

vision-based sensing or specialized geophysical instruments [38,39]. These methods have demonstrated high accuracy in detecting surface and subsurface pavement defects and have been widely adopted in intelligent infrastructure monitoring systems [40]. However, their practical deployment is often constrained by environmental conditions and hardware requirements. Vision-based approaches are highly sensitive to illumination conditions and may suffer performance degradation in nighttime or low-visibility environments. In addition, many sensor-based techniques require dedicated inspection vehicles and complex hardware configurations, which limit their scalability for continuous and large-scale pavement monitoring. Although tire–road coupling noise has been recognized as a potential information source for pavement condition assessment, its application in automated pavement inspection remains limited for the following reasons: (a) Computational frameworks for transforming tire–road acoustic signals into structured engineering information remain underdeveloped. While acoustic signals contain rich information on tire–pavement interaction, effective feature representation methods for pavement defect detection are still insufficiently explored. (b) There is a lack of integrated learning frameworks capable of modeling the temporal–spectral characteristics of tire–road acoustic signals. Existing studies typically rely on traditional signal processing or shallow learning models, which have limited capacity to capture complex acoustic patterns associated with pavement defects.

To address these gaps, this paper presents a vehicle-mounted system for collecting tire-road coupling noise under real urban driving conditions, and establishes an acoustic pavement defect dataset covering multiple defect types. A rapid pavement inspection method is further proposed for surface defect detection based on the established acoustic dataset. In addition, tire–road coupling noise is transformed into structured engineering information for pavement condition assessment. Acoustic signals generated from tire-road interaction are formalized as a computationally interpretable information source and integrated with MFCC-based feature representation and a CNN-SE-BiLSTM model, enabling robust and continuous pavement defect detection under real driving conditions.

2. Materials

2.1. Data collection equipment

Fig. 1 illustrates the vehicle-mounted system developed in this study for collecting tire-road coupling noise and corresponding pavement surface information. The system is installed on a standard urban passenger vehicle and integrates an audio acquisition module, an image acquisition module, a GPS module, and a laptop for data synchronization and storage. Tire-road coupling noise is captured using a DJI Mic Mini microphone with a wireless transmitter mounted near the rear tire. The microphone is positioned approximately 25 cm from the tire–road contact area and 40 cm above the ground to ensure consistent acoustic signal acquisition during vehicle motion. A foam windscreen is applied to reduce wind-induced noise during data acquisition. The transmitter wirelessly relays acoustic signals to a receiver connected to the laptop, enabling stable and high-quality real-time audio recording during vehicle operation. To associate the acoustic signals with pavement surface conditions, an Insta360 AcePro2 visual sensor is mounted on the same side of the vehicle and connected to the laptop. The captured images are saved for subsequent data annotation and validation. Meanwhile, spatial coordinates of each synchronized audio-image sample are recorded using a GPS system.

Table 2
Distribution of samples in the acoustic dataset.

Subset	Normal	Crack	Pothole	Total
Training	154	154	126	434
Validation	44	44	36	124
Testing	22	22	18	62
Total	220	220	180	620

2.2. Dataset establishment

To develop an acoustic pavement defect detection model, a tire-road coupling noise dataset was collected in Guangzhou, China. The selected road section has a total length of approximately 15 km. Three pavement surface conditions were considered, including AC-13 asphalt pavement, AC-16 asphalt pavement, and wet AC-13 asphalt pavement. During the data acquisition, the vehicle-mounted system maintained a driving speed of 30-40 km/h, with an ambient temperature of approximately 26 °C. The vehicle-mounted system was configured with an audio sampling rate of 48 kHz and a 24-bit sampling depth, while images were captured at a resolution of 8192 × 6744 pixels. In total, 35 min and 15 s of audio data and approximately 7,500 pavement images were collected. The dataset covers three pavement condition types, namely normal pavement, cracked pavement, and potholed pavement, as illustrated in Fig. 2. Audio, image, and GPS data are stored in WAV, JPG, and text formats, respectively. Audio segments corresponding to specific pavement conditions and longer than 0.35 s are manually labeled using the Audacity audio editor. Finally, the dataset comprises a total of 620 sample groups, including 220 groups of normal pavement, 220 groups of cracked pavement, and 180 groups of potholed pavement. The acoustic dataset is

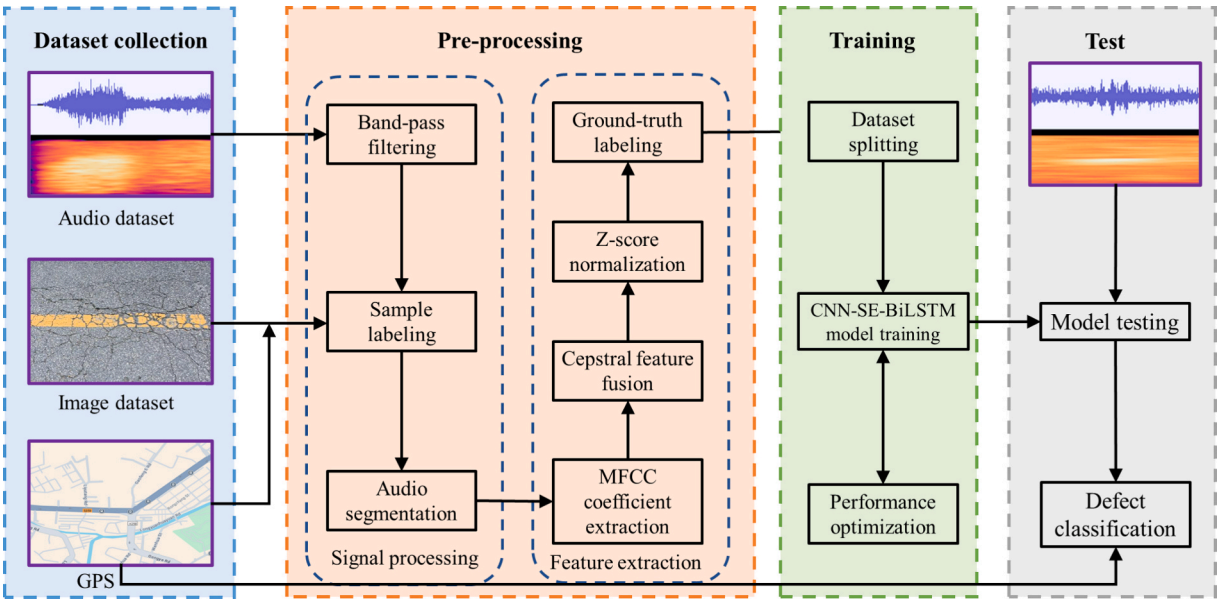
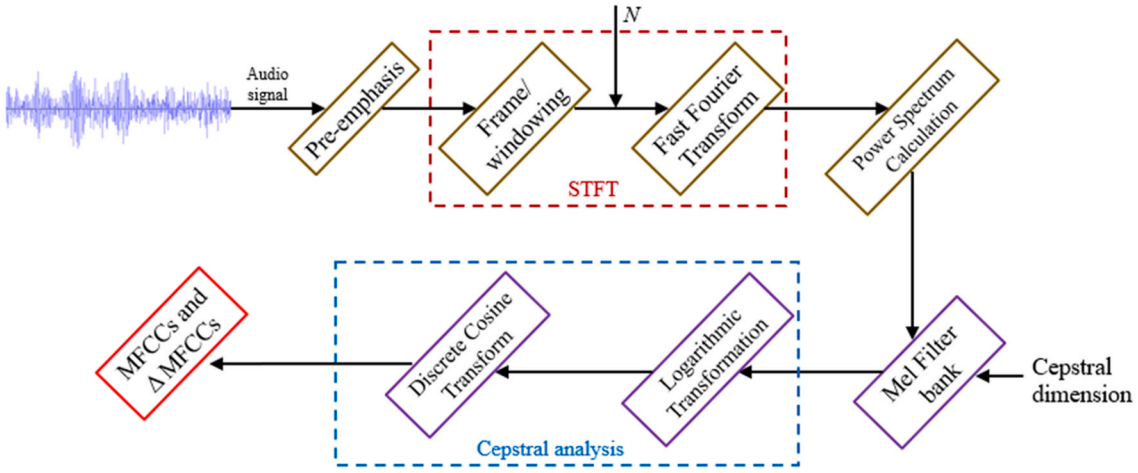


Fig. 3. Flowchart of the proposed pavement inspection model for road surface defects.



(a)

MFCC_0	MFCC_1	MFCC_2	...	ΔMFCC_0	ΔMFCC_1	...	ΔMFCC_12	Label
1.824953	0.291375	0.574219	...	0.324518	0.248375	...	0.175492	Normal
2.145672	1.238457	0.865324	...	0.315842	0.204765	...	0.287431	Pothole
0.947182	1.572864	0.678415	...	0.258431	0.142389	...	0.316542	Crack
0.415672	0.738254	0.214385	...	0.475821	0.135872	...	0.874512	Normal
1.578391	0.893214	0.472651	...	0.263482	0.193215	...	0.385127	Pothole
0.865273	1.342817	0.543672	...	0.314872	0.217584	...	0.251736	Crack
1.143275	0.753182	0.245691	...	0.294178	0.181273	...	0.462895	Normal
...	...	...	...	...	...	...	...	...

(b)

Fig. 4. (a) Workflow of MFCC feature extraction and (b) the corresponding MFCC feature matrix.

randomly divided into training, validation, and testing subsets with a ratio of 7:2:1, and the detailed sample distribution is summarized in Table 2.

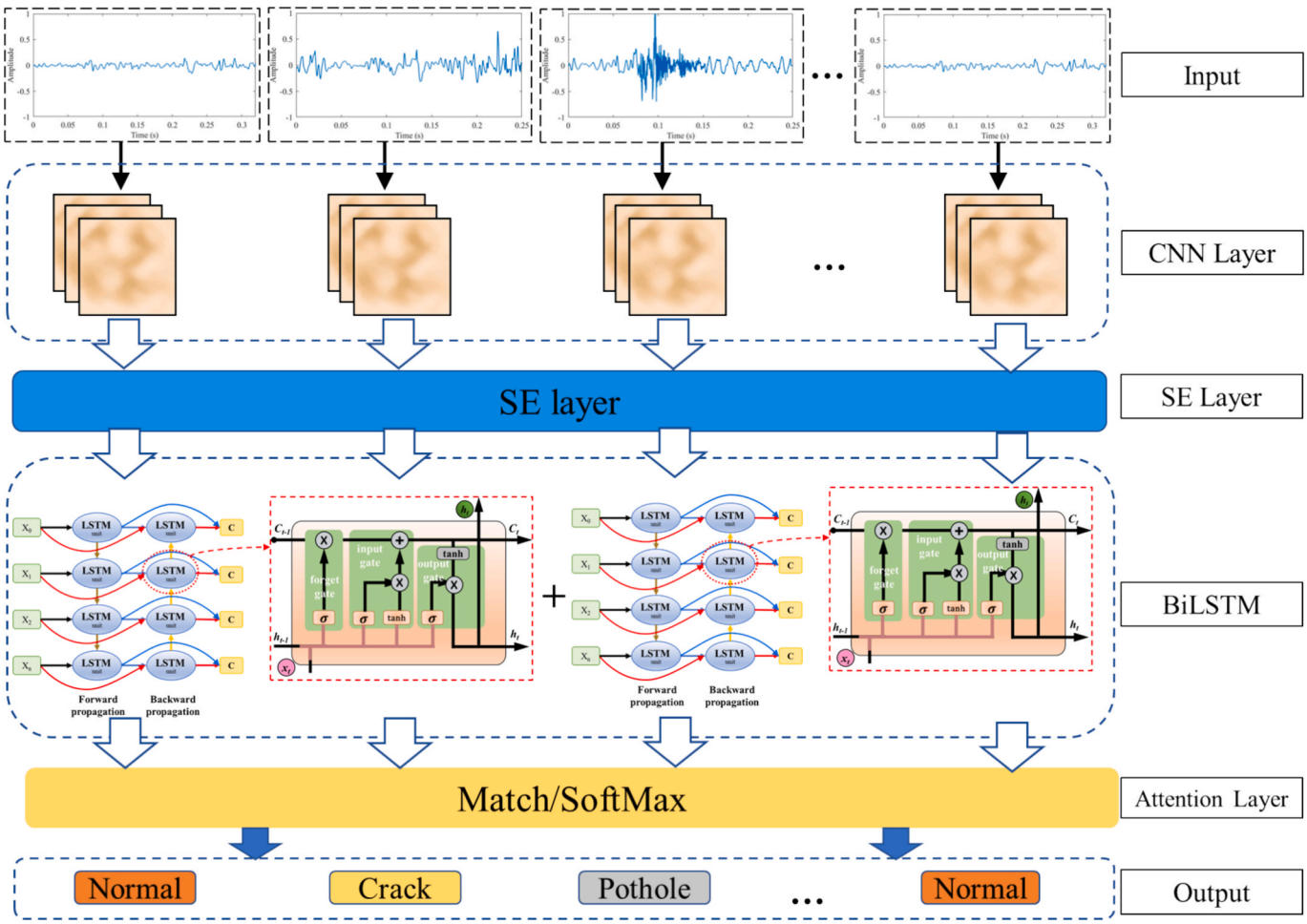


Fig. 5. Network architecture of the proposed CNN-SE-BiLSTM model.

3. Methodology

3.1. Data preprocessing

Fig. 3 illustrates the flowchart of the proposed pavement inspection model for road surface defects. The preprocessing of tire-road coupling signal consists of three main stages. Firstly, pavement condition labels are assigned to audio segments based on synchronized timestamps and GPS information. Then, each labeled audio segment is processed using short-time Fourier transform (STFT) and cepstral analysis to extract MFCC together with their first-order temporal derivatives (ΔMFCC). Finally, each audio segment is represented by a feature matrix of 13 MFCC and 13 ΔMFCC coefficients, with labels of normal pavement, cracked pavement, or potholed pavement.

3.2. MFCC feature extraction

MFCC provide a compact representation of acoustic spectra by mapping signal energy onto the mel scale, which reflects the nonlinear sensitivity of human auditory perception. This representation is widely used in acoustic analysis and has proven effective in characterizing spectral responses generated by tire-road coupling noise. The MFCC extraction process involves framing, windowing, Fourier transform, mel filterbank analysis, and cepstral transformation, resulting in coefficients that capture the dominant spectral characteristics of audio signals, as illustrated in Fig. 4. In this study, the MFCC extraction is configured with a frame length of 25 ms and a frame shift of 10 ms using a Hamming window. A total of 26 Mel filterbanks are applied to compute the cepstral coefficients, resulting in 13 MFCC features and their corresponding 13 ΔMFCC features. The first MFCC coefficient represents the overall spectral energy, while the remaining coefficients represent the energy distribution across mel frequency bands. The MFCC are sensitive to spectral variations induced by different pavement surface conditions. In addition, the ΔMFCC capture short-term spectral dynamics and highlight transient acoustic variations arising from localized surface irregularities.

Data annotation is performed by integrating GPS information with pavement images to accurately determine the spatial locations of audio segments and assign the corresponding pavement condition labels. Each audio segment is transformed into a feature matrix of dimension $n \times 26$, where n denotes the number of frames and the 26 columns correspond to MFCC and ΔMFCC features. To reduce scale differences and improve numerical stability during model training, Z-score normalization is applied to each feature dimension, and it is defined as follows:

$$Z = \frac{(X - \mu)}{\sigma} \tag{1}$$

where X denotes the original feature value, μ represents the mean, and σ is the standard deviation. This normalization converts the features into a standard normal distribution with zero mean and unit variance. The resulting normalized feature matrices are then used for classifying normal pavement, cracked pavement, and potholed pavement.

3.3. CNN-SE-BiLSTM model

To capture the temporal-spectral characteristics encoded in the

Table 3
Detailed architecture of the proposed CNN-SE-BiLSTM model.

Layer	Parameters	Output Description
Input	$T \times H \times W \times C$	Time × Height × Width × Channel
Conv + BN + ReLU	Kernel: 3×3×3, Filters: 32, Stride: (1,2,2)	Extracts low-level spatio-temporal features
SE Block	Reduction ratio r=8	Channel recalibration
MaxPool	Pool size (1,2,2)	Spatial downsampling
Conv + BN + ReLU	Kernel: 3×3×3, Filters: 64	Mid-level feature extraction
SE Block	r=8	Enhances important channels
MaxPool	Pool size (2,2,2)	Further spatio-temporal compression
Conv + BN + ReLU	Kernel: 3×3×3, Filters: 128	Deep spatio-temporal representation
SE Block	r=8	Adaptive channel weighting
Global Average Pooling (spatial)	/	Produces time-series features: $T/2 \times 128$
BiLSTM	Units:128, return_sequences=True	Temporal dependency modeling
BiLSTM	Units:64, return_sequences=False	Final temporal aggregation
Dropout	p=0.5	Regularization
Dense + ReLU	128 neurons	Fully connected layer
Softmax	num_classes	Classification output

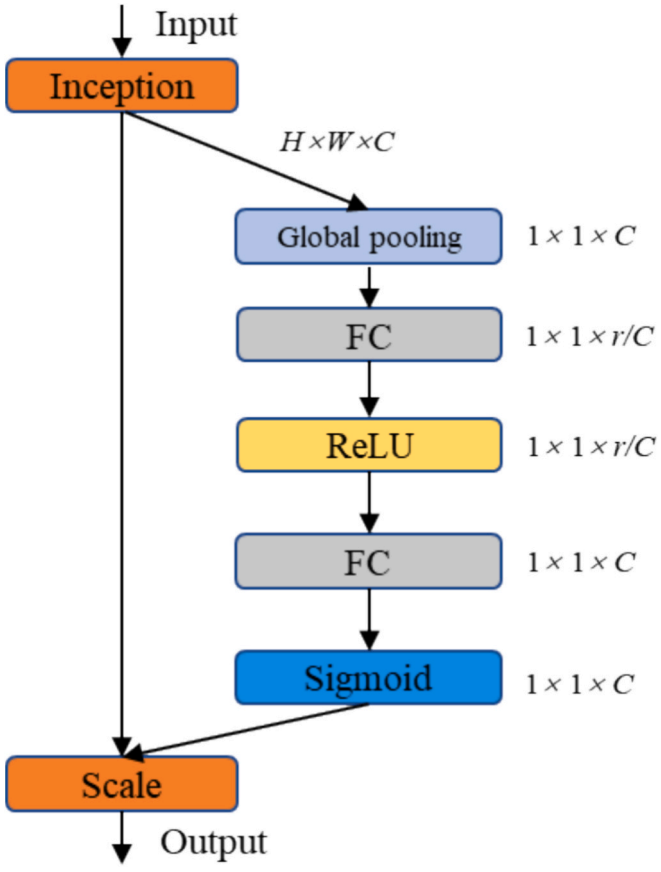


Fig. 6. Network architecture of the SE module.

MFCC feature matrix, a hybrid feature extraction and sequence modeling framework is constructed by integrating convolutional layers, squeeze-and-excitation (SE) channel attention [41], and a bidirectional long short-term memory (BiLSTM) module [42]. The proposed CNN-SE-BiLSTM model consists of three key components, where the CNN module extracts local temporal–spectral features, the SE module enhances channel-wise feature representation, and the BiLSTM module captures long-range temporal dependencies in the acoustic signal. The overall architecture of the proposed CNN-SE-BiLSTM model is illustrated in Fig. 5. A detailed specification of the model configuration, including layer types, kernel dimensions, activation functions, parameter settings, and output shapes, is provided in Table 3. The convolutional layers operate on the two-dimensional MFCC representation to learn localized temporal-spectral patterns, while the BiLSTM module models long-range temporal dependencies across consecutive frames. This combination enables effective discrimination of acoustic features across different pavement surface conditions.

3.3.1. SE module

Although convolutional layers are effective in extracting local temporal-spectral patterns from MFCC features, conventional convolution treats all feature channels equally and lacks the ability to adaptively emphasize informative frequency bands. To enhance channel-wise feature discrimination among MFCC and ΔMFCC coefficients, a SE attention module [41] is incorporated into the CNN feature extraction module. The SE module adaptively recalibrates channel responses by modeling the importance of each feature channel, strengthening informative spectral features related to pavement defects. The structure of the SE attention module is illustrated in Fig. 6. The SE block consists of two operations, squeeze and excitation. Given an input feature map $X_c \in \mathbb{R}^{C \times H \times W}$, where C, H, and W denote the number of channels, temporal length, and frequency dimension, respectively, the squeeze operation performs global average pooling to aggregate global contextual information into a channel descriptor:

$$Z_c = \frac{1}{HW} \sum_{i=1}^{H} \sum_{j=1}^{W} X_c(i,j) \tag{2}$$

where Z_c denotes the aggregated response of the c-th channel. This operation captures global temporal-spectral statistics and provides a compact descriptor of channel-level information. Then, the excitation operation models channel interdependencies by passing z through a bottleneck structure with a reduction ratio to generate channel weights:

$$s = \sigma(W_2 \delta(W_1 z)) \tag{3}$$

where W_1 and W_2 are parameter matrices of two fully connected layers, δ denotes the ReLU activation function, and σ represents the sigmoid function. This transformation captures nonlinear relationships among channels and computes a set of channel-wise scaling coefficients. Finally, the channel-wise attention weights are applied to the original feature map to obtain the recalibrated output:

$$Y_c = S_c \cdot X_c \tag{4}$$

where S_c denotes the excitation weight corresponding to the c-th channel. This operation adaptively emphasizes informative cepstral components while suppressing less relevant or noise-sensitive features within the MFCC representation. By integrating the SE mechanism into the CNN feature extraction module, the model is able to explicitly model channel-wise dependencies among MFCC and ΔMFCC coefficients, facilitating more effective feature representation for subsequent sequence modeling and classification tasks.

3.3.2. BiLSTM module

To enhance the temporal modeling capability of the proposed model, a BiLSTM module [42] is incorporated to capture bidirectional dependencies within the MFCC feature sequences, as illustrated in Fig. 7. Unlike the conventional LSTM module, the BiLSTM module models temporal dependencies in both forward and backward directions, thereby capturing global temporal context within an audio segment. This bidirectional processing is well suited for tire-road coupling signals,

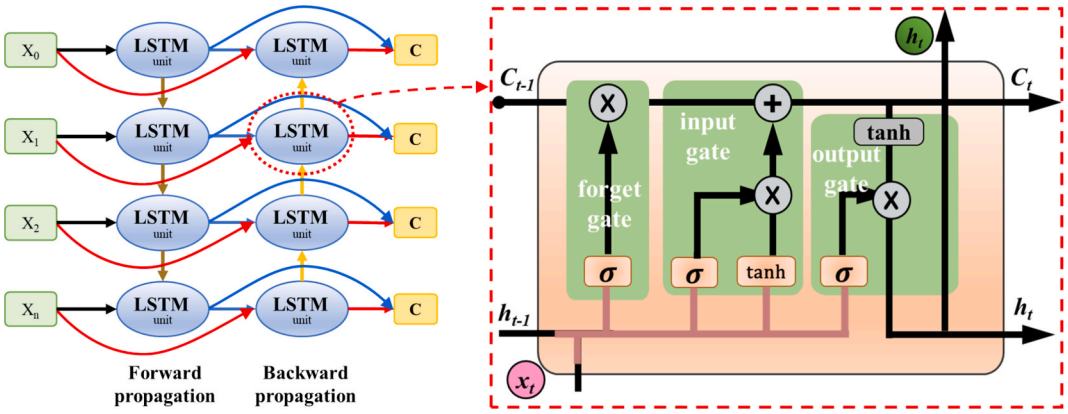


Fig. 7. Network architecture of the BiLSTM module.

where acoustic patterns of pavement defects exhibit temporal continuity and cross-frame correlations. Let $X = \{X_1, X_2, \cdots, X_n\}$ denote an MFCC feature sequence consisting of n frames. The forward and backward LSTM units generate hidden states $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$, respectively, which are defined as:

$$\overrightarrow{h_t} = \mathrm{LSTM}_f\left(x_n, \overrightarrow{h_{t-1}}\right) \tag{5}$$

$$\overleftarrow{h_t} = \mathrm{LSTM}_b\left(x_n, \overleftarrow{h_{t-1}}\right) \tag{6}$$

where LSTM_f and LSTM_b denote the forward and backward LSTM transformations. The output of the BiLSTM layer at each time step is obtained by concatenating the two directional hidden states:

$$h_t = \left[\overrightarrow{h_t}, \overleftarrow{h_t}\right] \tag{7}$$

This representation integrates temporal information from both past and future contexts, enabling the proposed model to capture structural variations in cepstral patterns across different pavement surface conditions. The BiLSTM layer is followed by a fully connected classifier, outputting the probabilities of normal, cracked, and potholed pavement. By incorporating the BiLSTM module, the model is able to preserve long-range temporal dependencies within MFCC sequences and provide a richer temporal representation for subsequent classification.

3.4. Evaluation indicators

All experiments were conducted on a desktop workstation running

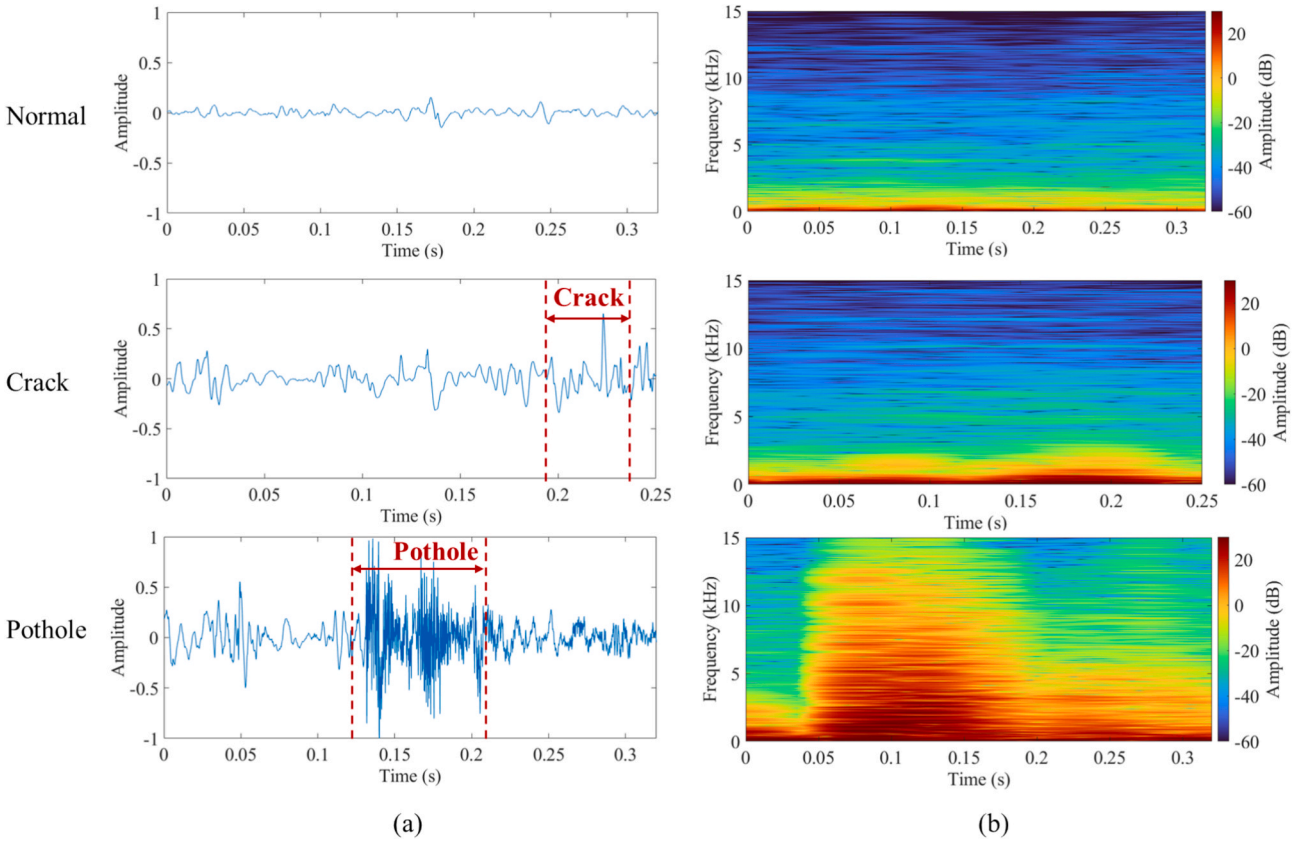


Fig. 8. Corresponding (a) audio waveforms and (b) spectrograms for three pavement conditions: normal, crack, and pothole.

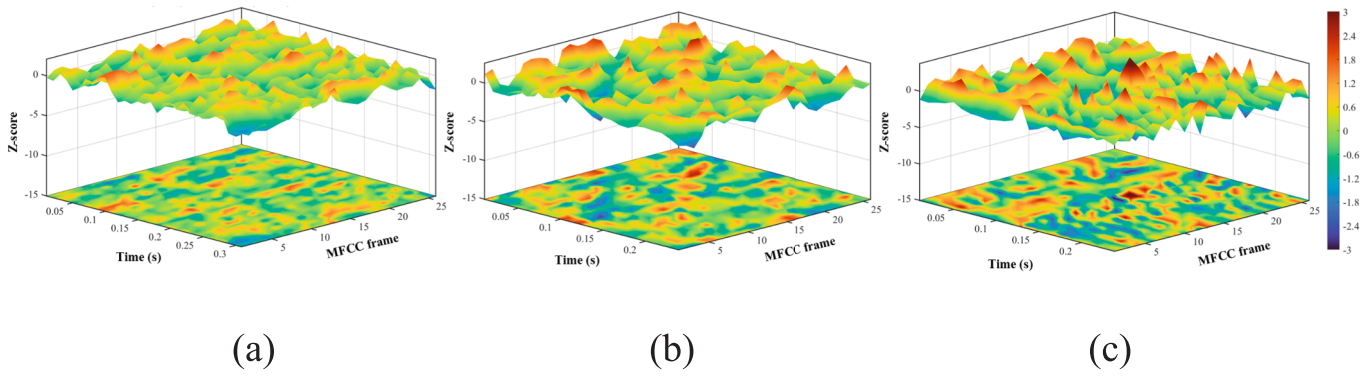


Fig. 9. Three-dimensional MFCC representations under different pavement conditions: (a) normal, (b) crack, and (c) pothole.

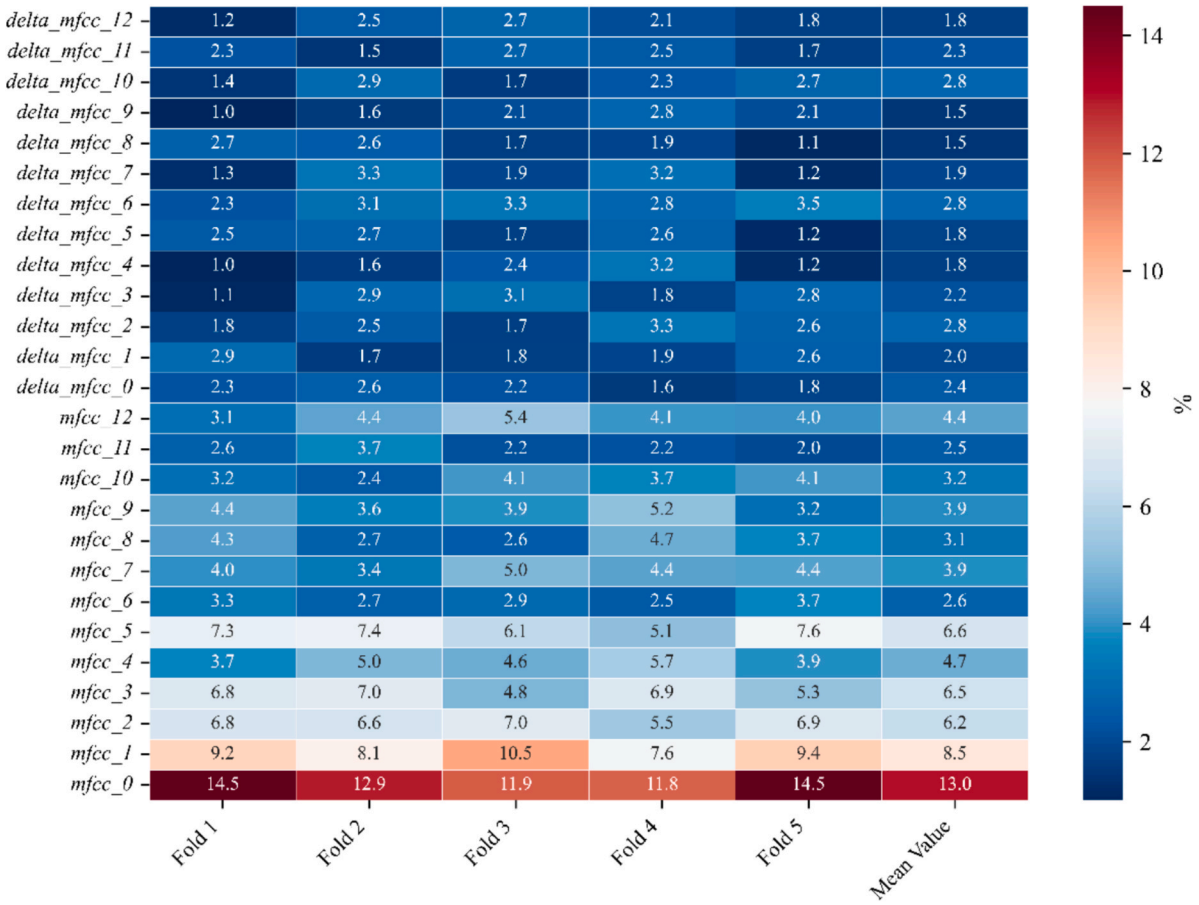


Fig. 10. Feature importance results obtained using five-fold cross-validation.

Windows 11 Professional, equipped with an AMD Ryzen 9 9950X 16-core processor and an NVIDIA GeForce RTX 4090D GPU with a VRAM of 24 GB. Model development and training were carried out using open-source frameworks, including PyTorch 2.3.0, CUDA 12.1, and Python 3.12.0. The performance of the detection model is evaluated using accuracy, precision, recall, and F1-score. These metrics provide a comprehensive assessment of the model's classification performance across different pavement defect types. To ensure reliable model evaluation and reduce the risk of overfitting, a five-fold cross-validation method [43] is adopted in this paper. The final performance metrics are obtained by averaging the results across all folds, providing a robust estimate of the model's generalization capability and reducing the influence of randomness arising from data splitting.

4. Model evaluation

4.1. Feature representation analysis

The three pavement conditions exhibit different time–frequency characteristics in both waveform and spectrogram representations, as shown in Fig. 8. For normal pavement, the audio waveform remains relatively stable with low-amplitude fluctuations, and the corresponding spectrogram shows a relatively uniform energy distribution. Cracked pavement presents more irregular amplitude variations in the waveform; however, its spectrogram patterns are not always characterized by clearly distinguishable frequency-band energy concentrations, since the induced acoustic disturbances are often weak and distributed. Potholes conditions produce a more pronounced acoustic response than cracked

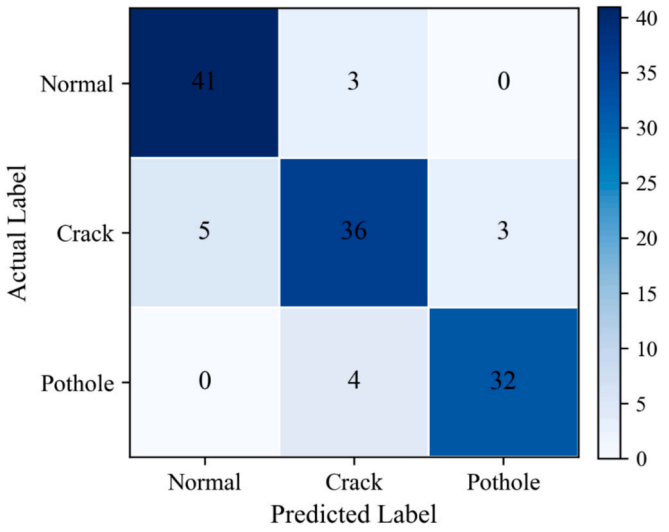


Fig. 11. Confusion matrix between predicted and true labels on the validation dataset for pavement defect detection.

pavement, characterized by sharp amplitude peaks in the waveform and intensified high-frequency components in the spectrogram. These observations indicate that while severe defects generate visually distinguishable time-frequency signatures, subtle defects such as cracks may not be reliably separated using spectrograms alone.

Fig. 9 presents three-dimensional representations of MFCC features corresponding to different pavement conditions. Normal pavement shows stable and uniform MFCC patterns, reflecting smoothness in tire-road interaction. In comparison, cracked pavement is characterized by localized increases in mid-order MFCC coefficients, indicative of transient acoustic variations arising from surface irregularities. Pothole conditions exhibit clear variability, characterized by abrupt coefficient fluctuations and increased dispersion across multiple MFCC dimensions. These varying temporal-cepstral patterns confirm the sensitivity of MFCC representations to pavement surface irregularities and provide informative features for subsequent model learning and classification.

4.2. Model performance evaluation

Fig. 10 presents the averaged importance of the 26 MFCC features obtained from five-fold cross-validation. The results indicate that the fundamental and mid-order cepstral coefficients contribute more prominently in characterizing pavement-related acoustic responses. Among these features, the fundamental coefficient mfcc_0 shows the highest average contribution (13.0%), followed by mfcc_1 (8.5%), mfcc_5 (6.6%), mfcc_3 (6.5%), and mfcc_2 (6.2%). These coefficients exhibit relatively consistent importance rankings across different folds, indicating stable contributions to pavement condition discrimination. In comparison, ΔMFCC features show lower importance values, ranging from approximately 1.0% to 3.5%, suggesting that first-order temporal derivatives provide complementary information to static cepstral features. Overall, feature importance analysis shows that pavement-related acoustic information is mainly captured by static MFCC features, while ΔMFCC features provide complementary information.

To further evaluate the detection performance of the proposed CNN-SE-BiLSTM model, a confusion matrix summarizing the relationship between predicted and true labels on the validation dataset is presented in Fig. 11. The results demonstrate high classification accuracy across all pavement types, with minimal confusion between normal and crack samples and clear separability of pothole samples. Representative classification results on the testing dataset are illustrated in Fig. 12. The corresponding precision values are 80.8% for normal pavement, 80.0% for cracked pavement, and 93.8% for potholed pavement, indicating that the proposed CNN-SE-BiLSTM model achieves reliable classification performance under the five-fold cross-validation method.

4.3. Robustness test

A comparative study is conducted to evaluate the effectiveness of different models for pavement defect detection, as summarized in Table 4. The LSTM model achieves an accuracy of 77.4%, while using bidirectional temporal modeling improves the accuracy to 80.6%. Further enhancement is obtained by incorporating the CNN module before the BiLSTM, leading to an accuracy of 82.3%, which confirms the benefit of convolution-based local feature representation. Among all models, the proposed CNN-SE-BiLSTM achieves the best performance, with an accuracy of 83.9% and corresponding precision, recall, and F1-score values of 84.9%, 83.8%, and 84.0%, respectively. These results confirm the effectiveness of combining convolutional feature extraction, SE attention, and bidirectional temporal modeling for pavement defect detection.

To further examine model performance under varying pavement

Table 4
Performance comparison of different models for pavement defect detection.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
LSTM	77.4	79.3	77.1	77.6
BiLSTM	80.6	82.0	80.5	80.9
CNN-BiLSTM	82.3	83.4	82.0	82.3
CNN-SE-BiLSTM	83.9	84.9	83.8	84.0

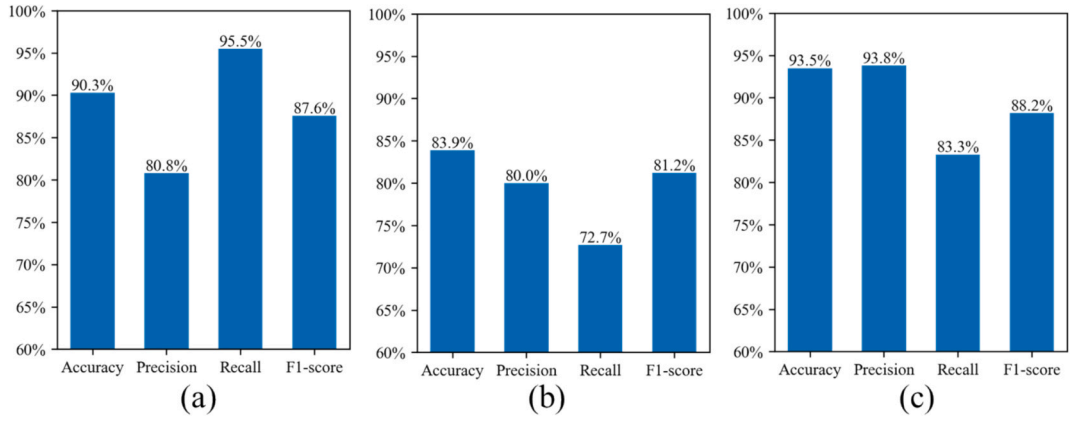


Fig. 12. Evaluation results on the testing dataset: (a) normal, (b) crack, and (c) pothole.

Table 5
Evaluation results of the proposed model under different pavement conditions.

Pavement Condition	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
AC-13 asphalt pavement	86.3	85.1	84.6	84.8
AC-16 asphalt pavement	87.1	86.0	85.4	85.7
Wet pavement (AC-13)	82.4	81.2	80.6	80.9
Wet pavement (AC-16)	80.7	79.5	78.9	79.2

materials and surface wetness conditions, Table 5 shows the classification results obtained for four pavement scenarios, including AC-13 asphalt pavement, AC-16 asphalt pavement, wet AC-13 pavement, and wet AC-16 pavement. It can be seen that the proposed model achieves accuracies of 86.3% and 87.1% on dry AC-13 and AC-16 pavements, respectively. Under wet conditions, the accuracy decreases to 82.4% for wet AC-13 pavement, indicating the tiny influence of surface wetness on acoustic characteristics. The model still maintains reasonable classification performance under this condition, achieving an accuracy of 80.7%, which can be attributed to the similar temporal–spectral characteristics shared between wet AC-16 pavement and other pavement conditions included in the training set. This result further demonstrates that the proposed acoustic-based approach can effectively capture invariant features of tire–road interaction without reliance on visual information, enabling robust performance under varying surface and environmental conditions.

5. Field experiment and results

A field experiment was conducted on an urban road in Guangzhou, China, to evaluate the performance of the proposed acoustic-based pavement inspection method under real driving conditions. Tire-road coupling noise was collected using the vehicle-mounted system while the test vehicle traveled along an approximately 2 km road segment at a speed of 30-40 km/h. Acoustic signals were recorded at a sampling rate of 48 kHz with a 24-bit resolution. A total of 12 pavement defects were collected, including 9 cracks and 3 potholes, based on field inspection records. The collected acoustic data were processed using the proposed CNN-SE-BiLSTM model to detect and classify pavement defect occurrences. Among the detected defects, four representative cases, consisting of 1 normal, 1 crack and 2 potholes, were selected for detailed analysis and visualization, as shown in Fig. 13. The results indicate that the proposed model successfully identify and classify the observed pavement defects under real driving conditions, with the predicted defect types consistent with field observations. These field experiment results demonstrate the effectiveness and practical applicability of the proposed acoustic-based inspection method for in-situ pavement defect detection.

From a practical deployment perspective, acoustic-based techniques provide an alternative sensing approach using tire–road coupling noise. Although the overall classification accuracy achieved in this study (83.9%) is lower than that reported for some image- and sensor-based

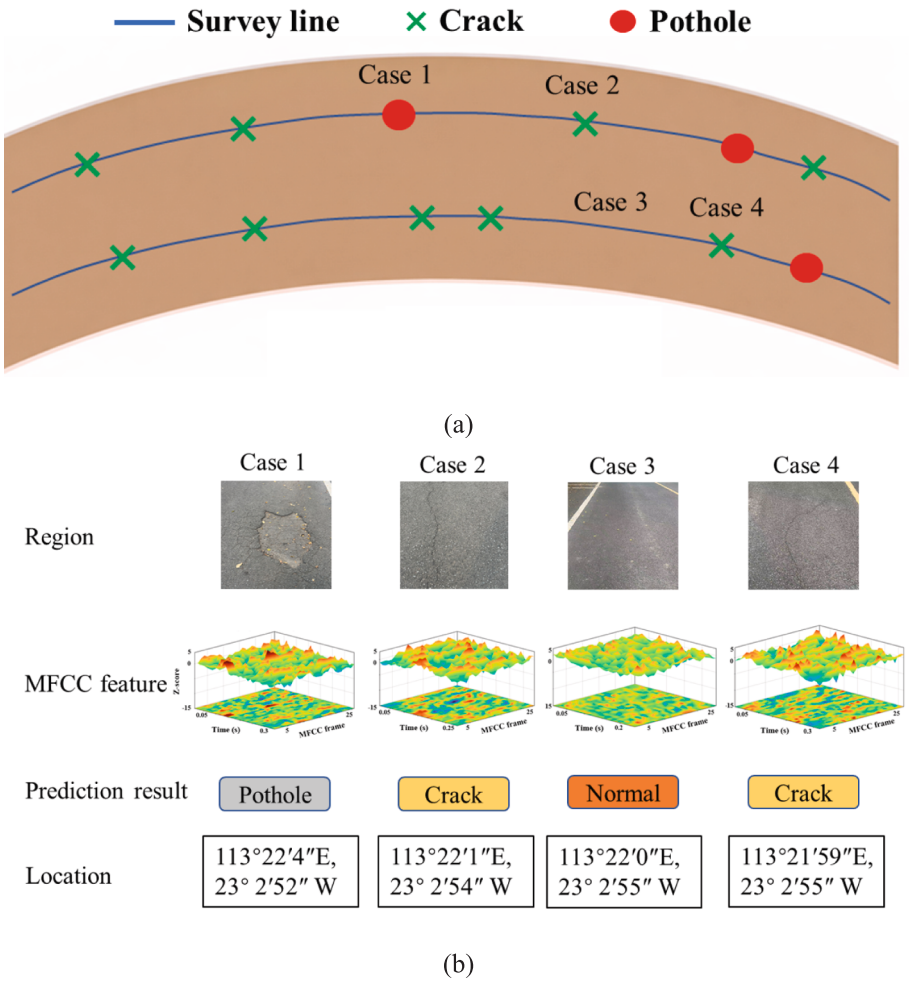


Fig. 13. Road surface detection on an urban road in Guangzhou, China. (a) Illustration of the inspection route and (b) the detected surface defects.

techniques, the proposed method offers advantages in deployment flexibility and inspection efficiency. Acoustic sensing enables continuous and non-contact inspection without reliance on controlled illumination conditions or specialized geophysical instruments. In addition, MFCC-based feature extraction requires substantially lower computational resources than high-resolution image processing or radar signal reconstruction. While acoustic sensing provides less precise defect localization than certain sensing modalities, it remains well suited for large-scale pavement screening due to its efficiency and scalability. Therefore, it can serve as a practical complementary solution for rapid and cost-effective pavement condition inspection. However, it should be noted that the number of detected defects in the field experiment is relatively limited, and future work will focus on large-scale validation across more diverse road conditions to further assess the generalization capability of the proposed method.

6. Conclusions

This paper presents a rapid and non-contact pavement defect inspection method based on tire-road coupling noise, with an aim to address the cost and deployment limitations of conventional vision-based techniques. A vehicle-mounted system was developed to enable continuous pavement condition inspection under real driving conditions. Then, MFCC are employed to characterize the temporal-spectral properties of tire-road coupling signals. A CNN-SE-BiLSTM model is designed to capture local time-frequency patterns, channel-wise feature relevance, and bidirectional temporal dependencies related to pavement surface defects. Experimental results indicate the proposed method achieves reliable classification performance, with an overall F1-score of 84.0% under a five-fold cross-validation method. The model exhibits stable performance across different pavement surface conditions. In addition, field experiments conducted on urban roads demonstrate that the proposed method is capable of identifying cracks and potholes from continuous acoustic measurements without reliance on visual sensing devices. These findings indicate that tire-road coupling noise enables efficient large-scale pavement condition inspection, particularly for rapid and low-cost deployment. From a practical perspective, the proposed method can be readily deployed on standard vehicles without requiring specialized inspection equipment, enabling continuous, large-scale, and cost-effective pavement monitoring in real-world urban environments.

CRediT authorship contribution statement

Yunpeng Yue: Writing – original draft, Methodology, Data curation, Conceptualization. **Hai Liu:** Writing – review & editing, Visualization, Project administration, Methodology. **Xiaoyu Liu:** Writing – review & editing, Validation, Investigation, Formal analysis. **Zhijie Chen:** Validation, Software, Methodology, Formal analysis. **Yi Li:** Writing – review & editing, Validation, Resources, Formal analysis. **Peng Lin:** Writing – review & editing, Resources, Formal analysis. **Jie Cui:** Supervision, Project administration, Methodology, Funding acquisition.

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.

Acknowledgments

This work was support by Guangdong Basic and Applied Basic Research Foundation (2024B1515040017, 2025A0505020086), National Natural Science Foundation of China (52379099), Science and Technology Projects of Guangzhou, China (2025B03J0132) and China Postdoctoral Science Foundation (No. 2024M760619).

Data availability

Data will be made available on request.

References

[1] J.S. Khichad, R.J. Vishwakarma, Overview and discussion of pavement performance prediction techniques for maintenance and rehabilitation decision-making, International Journal of Pavement Research and Technology (2024) 1–17.
[2] H. Li, R. Chen, N. Ritha, J. Wang, Z.a. Chen, An automatic detection method for road damage in UAV images based on multi-level perception and feature aggregation, Advanced Engineering Informatics, 69 (2026) 103814.
[3] X. Liu, K. Wu, X. Cai, W. Huang, Semi-supervised semantic segmentation using cross-consistency training for pavement crack detection, Road Materials and Pavement Design 25 (2024) 1368–1380.
[4] S. Shim, Semantic segmentation for crack detection via generative knowledge distillation, Automation in Construction 175 (2025).
[5] H. Zhang, A.A. Zhang, Y. Wei, J. Shang, Z. Dong, A. He, Robust variable resolution pixel-level detection of multiple distresses on rural roads, Advanced Engineering Informatics 69 (2026) 103976.
[6] G. Zhu, S.-L. Shen, J. Yao, M. Wang, J. Zhuang, Z. Fan, Automatic lightweight networks for real-time road crack detection with DPSO, Advanced Engineering Informatics 68 (2025) 103610.
[7] S. Jiang, H. Wang, Z. Ning, S. Li, Lightweight pruning model for road distress detection using unmanned aerial vehicles, Automation in Construction 168 (2024) 105789.
[8] W. Zhou, Y. Zhan, H. Zhang, L. Zhao, C. Wang, Road defect detection from on-board cameras with scarce and cross-domain data, Automation in Construction 144 (2022).
[9] A. Khan, S. Khan, M.A. Elhassan, H. Deng, Advancing road safety: A lightweight feature fusion model for robust road crack segmentation, Advanced Engineering Informatics 65 (2025) 103262.
[10] Z. Liu, X. Gu, W. Wu, X. Zou, Q. Dong, L. Wang, GPR-based detection of internal cracks in asphalt pavement: A combination method of DeepAugment data and object detection, Measurement 197 (2022) 111281.
[11] D. Zhang, X. Xu, H. Lin, R. Gui, M. Cao, L. He, Automatic road-marking detection and measurement from laser-scanning 3D profile data, Automation in Construction 108 (2019) 102957.
[12] X. Xiong, A. Meng, J. Lu, Y. Tan, B. Chen, J. Tang, C. Zhang, S. Xiao, J. Hu, Automatic detection and location of pavement internal distresses from ground penetrating radar images based on deep learning, Construction and Building Materials 411 (2024) 134483.
[13] H. Ge, J.C. Quezada, V. Le Houerou, C. Chazallon, P. Hornych, A new tire-sensor-pavement coupling chain for investigating asphalt mixture responses under rolling tire loads, Road Materials and Pavement Design 24 (2023) 245–262.
[14] S. Zhao, J. Zhang, Y. Jiang, C. He, J. Han, Tire-Road friction coefficients adaptive estimation through image and vehicle dynamics integration, Mechanical Systems and Signal Processing 224 (2025) 112039.
[15] D. Chen, N. Chen, X. Zhang, Y. Guan, Real-time road pothole mapping based on vibration analysis in smart city, IEEE Journal of selected topics in applied Earth observations and remote sensing 15 (2022) 6972–6984.
[16] P. Mikhailenko, Z. Piao, M.R. Kakar, M. Bueno, S. Athari, R. Pieren, K. Heutschi, L. Poulikakos, Low-Noise pavement technologies and evaluation techniques: a literature review, International Journal of Pavement Engineering 23 (2022) 1911–1934.
[17] K.R. Mannem, E. Mengiste, S. Hasan, B.G. de Soto, R. Sacks, Smart signal classification for tracking of construction tasks, Automation in Construction 165 (2024) 105485.
[18] S.-K. Lee, J. Yoo, C.-H. Lee, K. An, Y.-S. Yoon, J. Lee, G.-H. Yeom, S.-U. Hwang, Road type classification using deep learning for Tire-Pavement interaction noise data in autonomous driving vehicle, Applied Acoustics 212 (2023) 109597.
[19] C. Ramos-Romero, C. Asensio, R. Moreno, G. de Arcas, Urban road surface discrimination by tire-road noise analysis and data clustering, Sensors 22 (2022) 9686.
[20] S. Yang, Y. Chen, R. Shi, R. Wang, Y. Cao, J. Lu, A survey of intelligent tires for tire-road interaction recognition toward autonomous vehicles, IEEE Transactions on Intelligent Vehicles 7 (2022) 520–532.
[21] M.F. Cabral, F.S. Albuquerque, O. de Freitas Neto, T.M. Albuquerque, Study of Mechanisms That Generate Traffic Noise in the Tire–Pavement Interface for Low-Volume Roads in Brazil, Transportation Research Record 2474 (2015) 98–107.
[22] Z. Zhang, B. Luan, X. Liu, M. Zhang, Effects of surface texture on tire-pavement noise and skid resistance in long freeway tunnels: From field investigation to technical practice, Applied Acoustics 160 (2020) 107120.
[23] E. El-Din Hemdan, M.E. Al-Atroush, A Review Study of Intelligent Road Crack Detection: Algorithms and Systems, International Journal of Pavement Research and Technology (2025).
[24] E. Raslan, M.F. Alrahmawy, Y. Mohammed, A.S. Tolba, Evaluation of data representation techniques for vibration based road surface condition classification, Scientific reports 14 (2024) 11620.
[25] P. Kamranfar, D. Lattanzi, A. Shehu, S. Stoffels, Pavement distress recognition via wavelet-based clustering of smartphone accelerometer data, Journal of Computing in Civil Engineering 36 (2022) 04022007.

[26] A. Jawale, G. Magar, MFCC delta–delta energy feature extraction for clustering of road surface types, International Journal of Pavement Research and Technology 16 (2023) 631–646.
[27] F.S. Cabral, H. Fukai, S. Tamura, Feature extraction methods proposed for speech recognition are effective on road condition monitoring using smartphone inertial sensors, Sensors 19 (2019) 3481.
[28] L. De Donato, F. Flammini, S. Marrone, C. Mazzariello, R. Nardone, C. Sansone, V. Vittorini, A survey on -video based defect detection through deep learning in railway maintenance, IEEE Access 10 (2022) 65376–65400.
[29] K. Wang, X. Zhang, F. Wan, R. Chen, J. Zhang, J. Wang, Y. Yang, Wheel defect detection using attentive feature selection sequential network with multidimensional modeling of acoustic emission signals, IEEE Transactions on Instrumentation and Measurement 72 (2023) 1–14.
[30] J. Shen, N. Liu, H. Sun, S. Wu, Z. Liang, L. Han, Y. Zhang, D. Li, Lightweight Semantic Feature Extraction Model With Direction Awareness for Aerial Traffic Object Detection, IEEE Transactions on Intelligent Transportation Systems (2025).
[31] D. Li, Z. Jin, C. Guan, L. Ji, Y. Zhang, Z. Xu, J. Zhang, KACNet: Enhancing CNN feature representation with Kolmogorov-Arnold networks for medical image segmentation and classification, Information Sciences 122760 (2025).
[32] J. Shen, W. Zhou, N. Liu, H. Sun, D. Li, Y. Zhang, An anchor-free lightweight deep convolutional network for vehicle detection in aerial images, IEEE Transactions on Intelligent Transportation Systems 23 (2022) 24330–24342.
[33] J. Shen, N. Liu, C. Xu, H. Sun, Y. Xiao, D. Li, Y. Zhang, Finger vein recognition algorithm based on lightweight deep convolutional neural network, IEEE Transactions on Instrumentation and Measurement 71 (2021) 1–13.
[34] S. Farhadi, M. Corrado, G. Ventura, Automated acoustic event-based monitoring of prestressing tendons breakage in concrete bridges, Computer-Aided Civil and Infrastructure Engineering 39 (2024) 3700–3720.
[35] H. Behrooz, M. Ilbeigi, A. Rashidi, Real-time anomaly detection in construction equipment operations using unsupervised signal processing, Computer-Aided Civil and Infrastructure Engineering (2025).
[36] F. Alías, J.C. Socoró, X. Sevillano, A review of physical and perceptual feature extraction techniques for speech, music and environmental sounds, Applied Sciences 6 (2016) 143.
[37] P. Ananth, M. Kothandaraman, V.S. Ishwarya, Multi-Channel Enhancement using Dual-Stream Encoders with Attention Mechanisms and Spatial Discrimination GAN, Circuits, Systems, and Signal Processing (2025) 1–45.
[38] S. Bang, S. Park, H. Kim, H. Kim, Encoder–decoder network for pixel-level road crack detection in black-box images, Computer-Aided Civil and Infrastructure Engineering 34 (2019) 713–727.
[39] J.L. Porsani, E. Slob, R.S. Lima, D.N. Leite, Comparing detection and location performance of perpendicular and parallel broadside GPR antenna orientations, Journal of Applied Geophysics 70 (2010) 1–8.
[40] H. Liu, Y. Yue, C. Liu, B.F. Spencer, J. Cui, Automatic recognition and localization of underground pipelines in GPR B-scans using a deep learning model, Tunnelling and Underground Space Technology 134 (2023) 104861.
[41] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 7132–7141.
[42] P. Zhou, W. Shi, J. Tian, Z. Qi, B. Li, H. Hao, B. Xu, Attention-based bidirectional long short-term memory networks for relation classification, in: Proceedings of the 54th annual meeting of the association for computational linguistics (volume 2: Short papers), 2016, pp. 207–212.
[43] J.D. Rodriguez, A. Perez, J.A. Lozano, Sensitivity analysis of k-fold cross validation in prediction error estimation, IEEE transactions on pattern analysis and machine intelligence 32 (2009) 569–575.
[44] H. Zhang, J. Shang, Z. Dong, A. He, A.A. Zhang, Y. Liu, K.C. Wang, Z. Lin, Automated detection of pavement manhole on asphalt pavements with an improved YOLOX, Journal of Infrastructure Systems 29 (2023) 04023023.
[45] J. Zhai, Z. Sun, J. Huyan, W. Li, H. Yang, Feature representation improved Faster R-CNN model for high-efficiency pavement crack detection, Canadian Journal of Civil Engineering 50 (2022) 114–125.
[46] Z. Liu, D. Liu, L. Zhang, C. Cai, X. Peng, Real-time pavement distress detection based on deep learning and visual sensors, Road Materials and Pavement Design (2024) 1–19.
[47] Z. Liu, X. Gu, J. Chen, D. Wang, Y. Chen, L. Wang, Automatic recognition of pavement cracks from combined GPR B-scan and C-scan images using multiscale feature fusion deep neural networks, Automation in Construction 146 (2023) 104698.
[48] J. Jiang, P. Li, J. Wang, H. Chen, T. Zhang, Asphalt pavement crack detection based on infrared thermography and deep learning, International Journal of Pavement Engineering 25 (2024) 2295906.
[49] A. Karukayil, C. Quail, F.A. Cheein, Deep learning enhanced feature extraction of potholes using vision and LiDAR data for road maintenance, IEEE Access (2024).