Proceedings of the 4th International Conference on Mathematical Physics and Computational Simulation
DOI: 10.54254/2753-8818/2026.33081

Intelligent Identification and Graded Characterization of Hidden Defects in Highway Tunnel Linings Based on Multi-Source Nondestructive Testing Fusion

Jianlun Wang

Zhengzhou Advanced Research Institute of Harbin Institute of Technology, Zhengzhou, China
rara481846778@gmail.com

Abstract. The hidden flaws in the tunnel lines of the highways are hard to determine since their signals are small, overlapping and susceptible to environmental noise. Based on this study, a nondestructive testing system with multiple sources is offered to combine ground penetrating radar, infrared thermography and ultrasonic testing to intelligently detect and classify tunnel lining defects as graded. A common workflow is created, such as sensor alignment, signal preprocessing, feature construction, feature-level fusion and decision-level classification. Specimens of concrete with embedded voids, cracks and layer separation were prepared and repeated scanning was performed under both laboratory and semi-field conditions. Detection accuracy of 93.84 +1.27% was recorded with fused model which is 8.61 percentage points higher than the highest performing single source model. On four level defect grading, the proposed method had a macro-F1 of 0.912 +0.018 and minimized the confusion between neighboring severity classes. The findings indicate that multi-source fusion has the potential to offer consistent localization, accurate grading and greater robustness compared to the individual sensing techniques and can be used as a viable foundation of detailed assessment of tunnel lining.

Keywords: Tunnel lining defects, Multi-source NDT, Data fusion, Defect identification, Defect grading

1. Introduction

Highway tunnels linings are exposed to long term traffic loads, seepage of groundwater, thermal change, and aging of materials. In such combined actions, the hidden defects are slowly developing within the structure, such as air voids, inner cracks, debonding, and repair-layer detachment. Most of these defects cannot be seen on the surface during their initial stages, but can greatly affect stiffness, seepage routes and durability [1]. As soon as the growth of the defects achieves a critical size, leakage, spalling or local instability may take place. It is thus important to ensure proper identification of hidden defects in order to evaluate tunnel health and plan maintenance activities [2].

Only visual inspection does not give reliable data on hidden defects. Ground penetrating radar has a sensitivity to dielectric discontinuities and is commonly used to detect thickness variations and

DOI: 10.54254/2753-8818/2026.33081

void-induced reflections. Infrared thermography can be helpful in detecting shallow anomalies that alter local heat transfer behavior [3]. Ultrasonic testing records wave propagation, attenuation, and reflection in the concrete and interface areas. Yet, any specific modality can describe only a portion of the attributes of a defect. Radar could be affected by moisture and reinforcement; infrared behavior is highly sensitive to thermal conditions; ultrasonic signals are susceptible to scattering and quality of coupling materials.

The real challenge is thus the combination of heterogeneous signals based on different sensing principles. The data are also varying in their resolution and sampling density as well as their response to depth and geometry. At the same time, in engineering practice, the inspection of tunnels is not limited to binary defect detection. The severity grading is used in making maintenance decisions so that the most severe cases could be prioritized in repairs based on the structural risk. Specimen preparation, simultaneous acquisition, standard preprocessing, multi-level fusion modeling and performance assessment in detail are included in this work.

2. Literature review

2.1. Tunnel NDT techniques and applications

The tunnel inspection has been slowly changing over time to combine the use of more than a single sensor to observe it. The most popular application of ground penetrating radar is the determination of voids, thickness changes and concealed discontinuities. Shallow defect screening is commonly performed using infrared thermography, particularly when debonding or local thermal anomalies are present [4]. The role of ultrasonic testing as a method of assessing internal integrity based on the behavior of travel time and attenuation is still significant. In fact, any one of these methods can provide only a fragmented set of evidence, and in isolation, they tend to give unstable findings.

2.2. Data fusion techniques in structural monitoring

Fusion data processing in structural monitoring is often classified as feature level fusion, decision level fusion or hybrid fusion. The feature level fusion is the combination of the descriptors of several sensors prior to classification. Decision level fusion is the aggregation of the outputs of independent models using weighted averaging (or voting) [5]. The hybrid fusion tries to maintain complementarity without increasing redundancy. It is dependent on alignment precision, feature standardization and stability of classifiers.

2.3. Intelligent algorithms for defect detection

Smart algorithms are now being applied instead of manual threshold judgment to interpret defects. Support vectors machines, random forests, and gradient boosting are models that can learn non-linear decision boundaries based on structured feature sets. Their effectiveness remains quite sensitive to the quality of features as well as balance between classes. Graded characterization is a harder problem than binary detection in tunnel lining inspection since neighbouring defect classes can be very similar both in terms of geometry and signal response [6].

DOI: 10.54254/2753-8818/2026.33081

3. Methodology

3.1. Multi-source data acquisition setup

The experimental stand was based on the combination of ground penetrating radar, infrared thermography and ultrasonic testing on one scanning frame. The radar subsystem used 400 MHz, 600 MHz and 900 MHz antennas, and 600 MHz antenna was chosen as the main channel due to its balanced depth of penetration and spatial resolution. Every radar trace had 1024 points of sampling, and the distance between each line was 5 mm. The infrared subsystem consisted of a thermal camera of size 640 x 512 with a sensitivity of 0.04 °C. There were controlled heating of 120 s and then cooling of 300 s at 10 Hz. Pulse-echo testing with a center frequency of 1.0 MHz and a digitization rate of 10 MHz was used in the ultrasonic subsystem [7].

Each sensor was referenced to a common coordinate system that was marked on the surface of the specimen. Both radar and infrared and ultrasonic measurements were subsequently converted to a common 5 mm x 5 mm grid. The spatial drift after calibration was reduced to 1.8mm, so that cross-source feature construction remained consistent.

3.2. Data preprocessing and feature construction

The time-zero correction, background removal, gain adjustment, and band-pass filtering were used to process the radar signals [8]. Among the features that were extracted on the radargrams are peak amplitude, reflection energy, local entropy, instantaneous frequency, and curvature related descriptors. The infrared sequence was corrected by emissivity and smoothed over time and transformed into thermal contrasts compared to preheating. Its properties were temperature difference, cooling gradient and local texture index. The wavelet thresholding technique was applied to denoise ultrasonic signals and the following quantities were extracted: arrival time deviation, attenuation coefficient, reflected energy, and waveform kurtosis. The multimodal feature vector at every inspection point was defined as follows in Equation (1):

$$\mathbf{F} = \left[\mathbf{F}_{gpr}, \mathbf{F}_{irt}, \mathbf{F}_{ut}\right] \in \mathbb{R}^{d} \tag{1}$$

where $\mathbf{F}_{gpr}, \mathbf{F}_{irt}, \mathbf{F}_{ut}$ denote the radar, infrared, and ultrasonic feature subsets. All features were standardized by z-score normalization using only the statistics of the training set [9].

3.3. Multi-level data fusion modeling

The original 38-dimensional feature vector was reduced by principal component analysis to 17 dimensions while preserving 95.6% of cumulative variance. Three candidate classifiers were then trained on the reduced feature space: a radial-basis support vector machine, a random forest, and an extreme gradient boosting model. In parallel, source-specific classifiers were trained separately for each modality to support decision-level fusion.

The final probability of class c_k was obtained through weighted confidence aggregation as shown in Equation (2):

$$P\left(c_k \middle| \mathbf{x}\right) = \sum_{m=1}^{3} \omega_m P_m\left(c_k \middle| \mathbf{x}\right), \quad \sum_{m=1}^{3} \omega_m = 1 \tag{2}$$

DOI: 10.54254/2753-8818/2026.33081

where $P_m(c_k|\mathbf{x})$ is the posterior probability produced by modality m, and ω_m is the validation-based weight. The optimized weights were 0.39 for radar, 0.27 for infrared, and 0.34 for ultrasonic testing. Defects were finally divided into four severity grades according to the integrated size, depth, and signal response of the anomaly [10].

4. Experimental process

4.1. Test specimen preparation and defect configuration

The twelve reinforced concrete lining specimens were made with a size of 1200 mm x 800 mm x 300 mm. C35 is the concrete grade and the test value of 28-day compressive strength was 37.8 -/+ 1.6 MPa. Artificial flaws have been cast into the material. All in all, 48 hidden flaws were produced, i.e., 18 air voids, 16 planar separations, and 14 crack-like discontinuities. Widths of the defects varied between 5 mm and 30 mm, lengths - between 40 mm and 180 mm, and depth of burial - at 25 mm, 50 mm, 75 mm, and 100 mm. Out of those specimens, ten had steel reinforcements, and four had localized areas of moisture conditioning.

4.2. Data collection procedure and alignment

The scanning of each specimen was done in 21 parallel lines spaced by 40 mm. The measurements of radar, infrared and ultrasound were performed in sequence through the same coordinates. Temperature in the laboratory was fixed at 23.4 0.90C and relative humidity at 56.2 4.70C. Every sample was measured three times on different days. It resulted in 756 radar scan lines, 21,600 thermograms, and 9,072 ultrasonic A-scan registers. After edge exclusion and fine alignment correction, 18,432 valid fused sampling points were saved.

4.3. Model training and evaluation workflow

The data was split at the specimen level into training, validation, and test subsets by a ratio of 70:15:15, or 12,864, 2,764, and 2,804 points respectively. Borderline-SMOTE was used on the minority samples of Grade IV in the training set only, which grew to 1,884 points. Optimization of hyperparameters was done through grid search, and the random forest proved to have the best validation performance using 300 trees and a maximum depth of 14. The accuracy and precision, recall, F1-score, Cohen's kappa, AUC binary detection, and weighted mean absolute grade error of the grading were the metrics used to evaluate the model. The findings were all presented in terms of mean ± standard deviation across ten random seeds.

5. Results

5.1. Detection accuracy under different data sources

Table 1 presents the binary defect detection results. Among the single-source models, radar achieved the best overall performance with an accuracy of 85.23 ± 1.74% and an AUC of 0.901 ± 0.012. Ultrasonic testing achieved 83.71 ± 1.58% accuracy and showed relatively stronger precision for deeper defects beyond 75 mm, reaching 0.846 ± 0.021. Infrared thermography was highly responsive to shallow debonding, with a recall of 0.882 ± 0.019 for defects within 25 mm depth, but its overall stability decreased in moisture-conditioned specimens, where the accuracy dropped to 78.64 ± 2.03%.

DOI: 10.54254/2753-8818/2026.33081

The proposed hybrid fusion model achieved the best performance, with an accuracy of 93.84 ± 1.27%, precision of 0.931 ± 0.014, recall of 0.946 ± 0.011, F1-score of 0.938 ± 0.012, and AUC of 0.968 ± 0.008. Compared with the best single-source model, the absolute accuracy gain was 8.61 percentage points. In mixed-interference specimens containing both reinforcement and moisture zones, radar, infrared, and ultrasonic accuracies decreased to 80.42 ± 1.96%, 73.58 ± 2.31%, and 81.07 ± 1.88%, respectively, whereas the fusion model maintained 91.33 ± 1.41%. For small voids of 10 mm diameter at 75 mm depth, the false negative rate decreased from 0.218 ± 0.026 under radar-only detection to 0.094 ± 0.017 after fusion.

Table 1. Binary detection performance under different data sources

Model	Accuracy (%)	Precision	Recall	F1-score	AUC	Kappa
GPR only	85.23 ± 1.74	0.847 ± 0.018	0.861 ± 0.016	0.854 ± 0.015	0.901 ± 0.012	0.701 ± 0.023
IRT only	78.64 ± 2.03	0.781 ± 0.024	0.804 ± 0.021	0.792 ± 0.020	0.842 ± 0.017	0.569 ± 0.031
UT only	83.71 ± 1.58	0.832 ± 0.019	0.845 ± 0.018	0.838 ± 0.016	0.887 ± 0.013	0.669 ± 0.021
Feature-level fusion	91.62 ± 1.36	0.905 ± 0.016	0.922 ± 0.013	0.913 ± 0.014	0.953 ± 0.010	0.829 ± 0.018
Decision-level fusion	92.47 ± 1.29	0.917 ± 0.015	0.931 ± 0.012	0.924 ± 0.013	0.961 ± 0.009	0.846 ± 0.017
Proposed hybrid fusion	93.84 ± 1.27	0.931 ± 0.014	0.946 ± 0.011	0.938 ± 0.012	0.968 ± 0.008	0.874 ± 0.017

5.2. Performance of defect grading classification

Four-level severity characterization was more difficult than binary detection because Grade II and Grade III defects exhibited partially overlapping signal responses. The radar-only model obtained a macro-F1 score of 0.781 ± 0.026, while infrared thermography produced unstable results for buried severe defects, with Grade III and Grade IV recalls of 0.664 ± 0.041 and 0.611 ± 0.046. Ultrasonic testing was more robust for high-severity defects, giving a Grade IV recall of 0.801 ± 0.028, but it was less consistent for the lower grades.

The hybrid fusion model substantially improved grading performance. It achieved a macro-F1 score of 0.912 ± 0.018 and reduced the weighted mean absolute grade error to 0.214 ± 0.037. The misclassification rate between Grade II and Grade III fell from 17.8 ± 2.4% in the radar model and 19.6 ± 2.7% in the ultrasonic model to 8.3 ± 1.5% in the hybrid model. For Grade IV defects, the proposed method obtained a precision of 0.928 ± 0.019 and a recall of 0.917 ± 0.021. The mean predicted severity scores for true Grade I, II, III, and IV samples were 1.24 ± 0.18, 2.07 ± 0.23, 2.91 ± 0.27, and 3.76 ± 0.22, respectively, showing a clear ordered separation.

5.3. Comparative analysis of fusion strategies

Figure 1 compares the detailed grading performance of different strategies. Feature-level fusion improved classwise recall because it preserved complementary descriptors before classification, but some redundancy remained among correlated features. Decision-level fusion produced more stable outputs when one modality degraded. The hybrid strategy achieved the best overall performance by combining both routes. Ablation analysis showed that removing infrared features caused the largest reduction in Grade II recall, from 0.901 ± 0.020 to 0.843 ± 0.024, indicating that thermal diffusion information was critical for shallow interface-related defects. Removing ultrasonic features mainly reduced Grade IV precision from 0.928 ± 0.019 to 0.871 ± 0.023. Removing radar information caused the largest drop in overall macro-F1, from 0.912 ± 0.018 to 0.851 ± 0.022. In the semi-field

DOI: 10.54254/2753-8818/2026.33081

tunnel section, the hybrid model identified 11 of 13 anomalous regions confirmed by follow-up records, with an average anomaly confidence of 0.884 ± 0.067 in verified damaged zones.

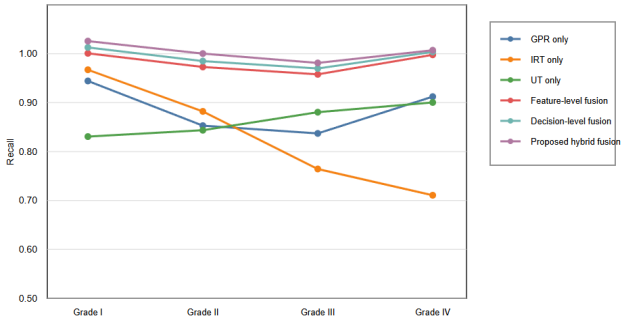

Figure 1. Four-level defect grading recall across different strategies

6. Conclusion

This paper has constructed an extensive framework of smart identification and graded description of hidden flaws in highway tunnel linings using multi-source nondestructive testing fusion. The approach combined synchronized acquisition, standardized preprocessing, feature construction, dimensionality reduction and weighted decision aggregation. As the findings indicated, the hybrid model suggested was more effective than any of the single-source baselines in the binary detection as well as four-level grading. It was found to have a detection rate of 93.84 ± 1.27 percent and a macro-F1 score of 0.912±0.018 on severity classification with no change on exposure to reinforcement and moisture interference. These findings suggest that fusion of multiple sources would offer more dependable and finer support of tunnel lining inspection compared to isolated sensing techniques.

References

[1] Liu, B., Zhang, J., Lei, M., & Yang, S. (2023). Simultaneous tunnel defects and lining thickness identification based on multi-task deep neural network from ground penetrating radar images. Automation in Construction, 145, 104633.
[2] Zhou, D., Zhang, Y., Li, H., & Chen, X. (2023). 3D visualization of karst caves in tunnels based on ground penetrating radar attribute analysis. Chinese Journal of Rock Mechanics and Engineering, 45, 310–317.
[3] Zhang, W., Li, Y., Wang, H., & Chen, Z. (2025). Vision and deep learning-based tunnel surface defect detection: A review. Measurement, 240, 114345.
[4] Zhang, X., Wang, Q., Liu, Y., & Sun, J. (2025). A scoping review of deep learning in non-destructive testing for construction. Electronics, 14(6), 1124.
[5] Poncetti, B. L., Rossi, M., & Ferrante, C. (2025). Tunnel inspection review: Normative practices and non-destructive testing methods. Infrastructures, 6(2), 41.
[6] Wang, Y., Li, X., Zhao, K., & Chen, L. (2025). Void detection behind tunnel concrete lining using ultrasonic-assisted ground penetrating radar method. Journal of Applied Geophysics.
[7] Qian, W., Liu, H., Zhang, Y., & Zhao, F. (2026). Non-destructive detection and three-dimensional imaging of underground defects using ground penetrating radar-based methods. Heritage Science.
[8] Luca, C. (2025). Advances in ground penetrating radar technology for non-destructive testing of concrete structures. Construction and Building Materials.
[9] Shimky, S., Mim, M., & Yao, F. (2025). Advances and challenges in non-destructive testing methods for concrete structures. World Journal of Engineering and Technology, 13, 791–815. https: //doi.org/10.4236/wjet.2025.134050
[10] Yao, Y., Chen, J., Liu, Z., & Wang, T. (2025). Intelligent tunnel lining defect detection: A comprehensive review of image acquisition and AI-based analysis. Infrastructure Intelligence Journal.