Data Engineering and Applications

Research Paper

Roadside Acoustic Array Data-Driven S4D-Transformer Model for Performance Evaluation of Noise Barriers

Sultan Al-Dhaheri[1, *], Khamis Al-Zahrani[1] and Hazaa Al-Mansoori[1]

[1] School of Engineering, Computer Science Department, Khalifa University, Abu Dhabi, PO Box 127788, United Arab Emirates
*Corresponding author: sultan.dh@kustar.ac.ae

Abstract. Most roadside noise barriers are still assessed through short-term manual surveys, so their actual noise-reduction performance under changing traffic composition, source height, pavement state, wind condition and receiver-side propagation environment remains insufficiently characterized. This paper proposes an S4D-Transformer model for data-driven performance evaluation using synchronous roadside acoustic-array recordings collected on the source side and protected receiver side of a full-scale barrier. First, multiple acoustic signals are converted into standardised time-frequency descriptors, spatial-coherence indicators, beam-direction features and traffic-state variables. A diagonal-structured state-space block is then employed to retain long-range acoustic memory across monitoring windows, and a transformer fusion layer estimates broadband insertion loss, spectral attenuation and rating confidence. A 28-day field dataset of 1,344 ten-minute windows and 7.8 million valid array frames was used for training and testing. The mean absolute insertion-loss error of the proposed model was 0.84 dB, the root-mean-square error was 1.12 dB, and the Pearson correlation coefficient with reference measurements was 0.946. Compared with CNN-LSTM, vanilla Transformer and pure S4D baselines, it reduced the mid-frequency error by 18.6%-31.4% and maintained 95% of the window-level uncertainty intervals within 2.7 dB. According to the above results, roadside acoustic arrays can perform non-intrusive, frequency-resolved and continuous barrier-performance diagnosis for operation and maintenance in complex urban corridors with minimal disturbance to normal traffic flow.

Keywords: *Roadside Acoustic Array, Noise Barrier Evaluation, S4D-Transformer, Insertion Loss, Spectral Attenuation*

Received on 13 March 2026, Accepted on 23 June 2026, Published on 05 July 2026

Introduction

Urban road corridors are continuous sources of the environment-noise problem because vehicle powertrain, tire-road interaction, pavement texture and traffic intermittency collectively generate sound exposure at adjacent receptors [1]. Traditional road-noise reduction methods have generally been fixed or absorbent barriers, but the acoustic advantage provided by a barrier rarely remains consistent over time after installation. Settlement, panel leakage, vegetation growth, drainage openings, surface contamination and changes in the proportion of heavy-vehicle traffic can all affect the effective insertion loss at the receiving end [2]. Measurement Standards provide the necessary basis for standardized tests, but the campaign-style protocols often describe only a small number of traffic and weather conditions. A barrier meets the acceptance criteria for a relatively dense road network but has poor performance under peak flow, wet pavement or high-wind conditions. Thus, the engineering question has moved from whether a barrier can be rated in an isolated survey to whether its performance can be continuously inferred from real roadside evidence. Acoustic arrays are suitable for this change because they can obtain directionality, coherence and source-receiver contrast, etc., instead of just a scalar sound-pressure level. Galiana Nieves et al. [3] assessed the sound reduction index of noise barriers with low sound insulation, showing why barrier performance should be evaluated as a configurationdependent acoustic response. Frequency-dependent attenuation is relatively small at low, mid- and high frequencies due to different diffraction and absorption mechanisms [4]. Field investigations have also shown that the range of receiver-side benefits for passenger-car-dominated and truck-dominated periods varies significantly [5]. The above observations motivate a model that treats barrier evaluation as a structured, time-varying acoustic inference problem rather than a single-number measurement task [6].

With the spread of low-cost microphones, embedded processors and extended-duration environmental monitoring, data-driven acoustic monitoring has been developing rapidly. Early learning models for traffic-noise prediction generally employed equivalent sound levels, traffic counts and basic weather information as inputs [7]. Recently, some research has used spectrograms, beamforming maps and recurrent networks to exploit the temporal variations in traffic scenes [8]. Barrier-Performance Evaluation is not the same as Traffic-Noise Prediction. The goal is a differential and physically constrained quantity: the reduction in the difference of the incident field caused by a shielding structure. A model that only observes the receiver end cannot distinguish between a quiet traffic period and high barrier effectiveness. If only short frames are taken, a slow change in lane occupancy or vehicle composition will not be shown. Transformers can model long-range dependencies, but their quadratic attention cost is inconvenient for many-channel acoustic streams sampled over extended periods [9]. State-space sequence models are also feasible options that can carry forward extended-range temporal information with a linear-complexity algorithm [10]. The S4 family has gained attention for modelling long signals in audio and sensor systems by embedding continuous-time dynamics into learnable discrete filters [11]. A diagonal S4D formulation is suitable for engineering deployment as it retains the long-memory property while reducing parameters and numerical complexity [12].

The current work introduces a roadside acoustic-array data-driven S4D-Transformer model for noise-barrier performance assessment. The first contribution is the use of an S4D block to retain extended acoustic context within a ten-minute monitoring window; the second is transformer-based fusion of spatial channels, spectral bands and traffic descriptors that are locally informative for barrier attenuation. The model estimates insertion loss, octave-band attenuation and a confidence score from synchronous source-side and receiver-side arrays. It is to be used in field research where traffic cannot be regulated and only reference measurements are available during calibration. In addition to a purely predictive road-noise model, the proposed method explicitly incorporates the source-receiver difference, array coherence and barrier-related frequency response before learning the evaluation mapping. This design can identify the following performance deficiencies in the model: low-frequency diffraction leakage, mid-frequency panel discontinuity, and high-frequency absorption attenuation.

The remainder of this paper is organized into five sections. Chapter II: Review of acoustic-array evaluation for road-noise barriers and learning-based sequence models in environmental acoustics. Chapter III shows the monitoring system, feature construction, S4D-Transformer architecture and objective functions used for performance evaluation. Chapter IV shows the field data, baseline comparisons, ablation analysis, frequency-resolved evaluation and engineering interpretation of the learned indicators. Chapter V is the conclusion of this paper, which summarizes the results, deficiencies and future work.

Related Work

Acoustic-array assessment of roadside noise barriers

In the past, most evaluations of roadside barriers have employed paired measurements or controlled-source tests and before-and-after comparisons. The above methods provide traceable indicators but are unable to cover the variations in active road corridors. Microphone arrays increase the range of measurement by adding space selection. Beamforming can locate the main traffic sources and reduce interference from other background noise [13]. Near-field array layouts can also be used to compare the incident and transmitted acoustic fields without assuming a uniformly spaced source line [14]. Barrier assessment is complicated because diffraction at the top edge, scattering at panel joints and leakage through openings may produce different spatial signatures at the receiver end. Previous array studies have indicated that coherence among channels can determine whether a measured value is mainly due to direct traffic, diffracted energy or local reflections [15]. The same information can help separate barrier-related attenuation from other traffic fluctuations.

Several engineering studies have further indicated that insertion loss is a spectral quantity. Low-frequency sound bends more easily around an obstacle, but high-frequency sound is more severely attenuated by material absorption, surface roughness and atmospheric absorption [16]. Therefore, if the purpose is only diagnosis and not regulatory reporting, octave-band or one-third-octave-band analysis should be used. Field research has shown that both the composition of traffic in the effective source height and the directivity change when heavy vehicles are in the outer lane or accelerating near the ramp [17]. A fixed measurement window may therefore overestimate or underestimate the barrier benefit if the traffic state is not representative. Although array-based tools have improved the observability of these mechanisms, many of the reported workflows still use manually selected intervals and deterministic post-processing [18]. There is a method deficiency for continuous evaluation in which the array signal is high-dimensional, the traffic state is nonstationary, and the desired output is a reliable performance index rather than a beamforming image.

Sequence learning models for environmental acoustic monitoring

Learning-based environmental-acoustic models have developed from shallow regressors at equivalent levels to neural networks that can process spectrograms and multi-sensor features. Convolutional networks are suitable for local time-frequency patterns of impulsive events and spectral slopes [19]. Recurrent networks can address the problem of time dependence, but training is difficult when there are variations in acoustic conditions over a long period or in synchronization of multiple channels [20]. Transformers address the problem of long-range dependencies through self-attention and have been applied to sound-event detection, acoustic scene classification, and traffic-flow inference [21]. Flexibility can be used to conduct a barrier test by examining attenuation changes at the source, receiver, frequency band and array geometry.

The computation cost of standard attention has promoted the application of state-space models to long sensor streams. Gu et al. [22] showed that structured state-space sequence layers can approximate continuous-time dynamics with learnable filters and handle long sequences at a linear or near-linear computational cost. Diagonal S4D layers reduce parameterization while maintaining stable memory, and are thus suitable for acoustic monitoring windows with slow traffic-state drift and fast pass-by events. Hybrid state-space and attention models have been recently introduced for audio and industrial sensing because the state-space part incorporates history, and attention focuses on informative channels or events [23]. Road-noise barrier literature has not yet proposed an architecture that links array-derived source-receiver descriptors with frequency-resolved performance evaluation. An S4D-Transformer model for barrier insertion loss, spectral attenuation and engineering confidence has been developed in this paper.

S4D-Transformer Evaluation Framework

Roadside array configuration and feature construction

The two linear microphone arrays in the monitoring system are synchronously deployed at the traffic side and the protected side of a roadside noise barrier. The source-side array represents the incident acoustic field, and the receiver-side array records the transmitted and diffracted field after barrier interaction. There are 8 calibrated microphones in each array, spaced 0.35m apart, and a sampling rate of 16kHz is used. A ten-minute window is divided into overlapping 1.0 s frames with a 50% hop, and a sequence is generated that is long enough to cover traffic-state changes but short enough for inspection-level reporting. Filter out saturation, rain effects and clock drift from the raw channels before building features. The retained frames are converted into octave-band energy, inter-channel coherence, source-receiver level difference, beam-direction entropy, and traffic descriptors estimated from acoustic pass-by counts.

The calibrated pressure of microphone m for a frame indexed by t is denoted as $x_m(t)$. The short-time spectral energy of band b is obtained by microphone sensitivity correction of the array-averaged spectrum. The input vector z_t contains the following: incident energy, receiver energy, coherence, directionality and traffic-state. A suitable physical contrast for barrier evaluation is used, and both the sound field before the barrier and the shielded field in the model are included in the vector. Figure 1 is the proposed roadside sensing workflow after synchronised acquisition and before performance output. The figure is designed to be a flow diagram because the engineering contribution depends on the connection among measurement, preprocessing, sequence modelling and rating.

https://doi.org/10.64972/dea.2026.v5i3.3961d:1-14

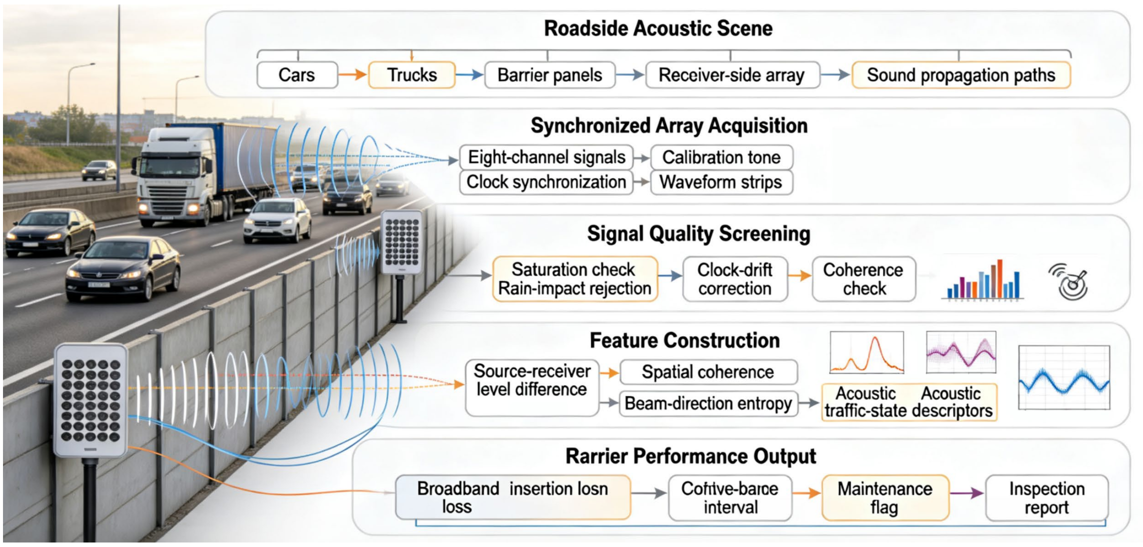

Figure 1. Roadside acoustic-array workflow for continuous noise-barrier performance evaluation.

The Feature window was intentionally set to the inspection scale and not at the acoustic-event scale. A single vehicle pass-by can show the source direction, but it is not suitable for describing the actual performance experienced by a residential receiver under normal road conditions. Alternatively, within the one-hour window, there may be a small dip in barrier performance amid a general rise. The ten-minute unit here balances the above two demands. It has sufficient diversity for the pass-by method of source normalization and can still assign underperforming intervals to a practical maintenance time stamp. The source-side and receiver-side descriptors in the feature tensor are not subtracted at the raw-signal level. First, these are calibrated and band-aligned to ensure coherence; otherwise, the later model would see uncontrolled microphone-level differences in the evidence.

The feature vector for each monitoring frame is defined as:

$$Z_t = [E_t^{src}, E_t^{rec}, C_t, H_t, Q_t] \qquad \text{Eq.(1)}$$

The objective indicator is the window-level insertion loss y_t, defined as the source-normalized reduction between the incident and receiver-side equivalent levels. A positive value indicates favourable shielding after calibration for the source-side variation. The same formulation is extended to octave bands, and then it can be determined whether a barrier is weak in the diffraction-dominated low band, the panel-transmission mid band, or the absorption-sensitive high band.

The octave-band insertion loss used as the main performance indicator is calculated as:

$$IL_b(t) = L_{eq,b}^{src}(t) - L_{eq,b}^{rec}(t) + \Delta_b \qquad \text{Eq.(2)}$$

S4D Temporal Encoder and Transformer Fusion

An S4D layer is used to build a sequence encoder that can map an acoustic-feature sequence to a latent state with stable long-term memory. Each latent mode of the diagonal state matrix will have a different decay speed and oscillatory characteristics. This is suitable for roadside acoustics as some characteristics change slowly with traffic density, and pass-by peaks and short leakage events occur over only a few seconds. The S4D layer processes the entire ten-minute window before applying attention to prevent excessively high weights from being given to isolated high-level events without considering the surrounding acoustic context.

The S4D encoder has the advantage of treating the acoustic record as an ordered physical process in practice. The model does not need to store all the raw frames; instead, it learns a small number of modes that can respond to repeated traffic patterns, slow background drift and short transient phenomena. This is not the same as a pure convolutional stack; the receptive field does not need to be expanded by depth or dilation, and long-range dependencies are not indirect. This is also different from using full self-attention on all frames, as the computational load is relatively high, and the model may overfit to repeated high-energy passages. In the proposed architecture, the state-space representation serves as the memory foundation, and attention is used for cross-feature fusion after the long sequence has been compressed.

The S4D temporal transition used to encode long acoustic memory is expressed as:

$$h_{t+1} = \bar{A}h_t + \bar{B}z_t, u_t = \bar{C}h_t + Dz_t \quad \text{Eq.(3)}$$

After S4D encoding, a transformer fusion layer merges the latent temporal state with band and channel embeddings. Multi-head attention only uses compact tokens derived from the state-space encoder and not raw high-rate frames. The design reduces computational cost for long observation windows because attention is applied to compact state tokens while still focusing on intervals with significant changes in barrier behavior. Figure 2 is the S4D-Transformer architecture, which includes an acoustic feature tokenizer, a diagonal state-space memory, transformer fusion and three output heads for insertion loss, spectral attenuation and confidence.

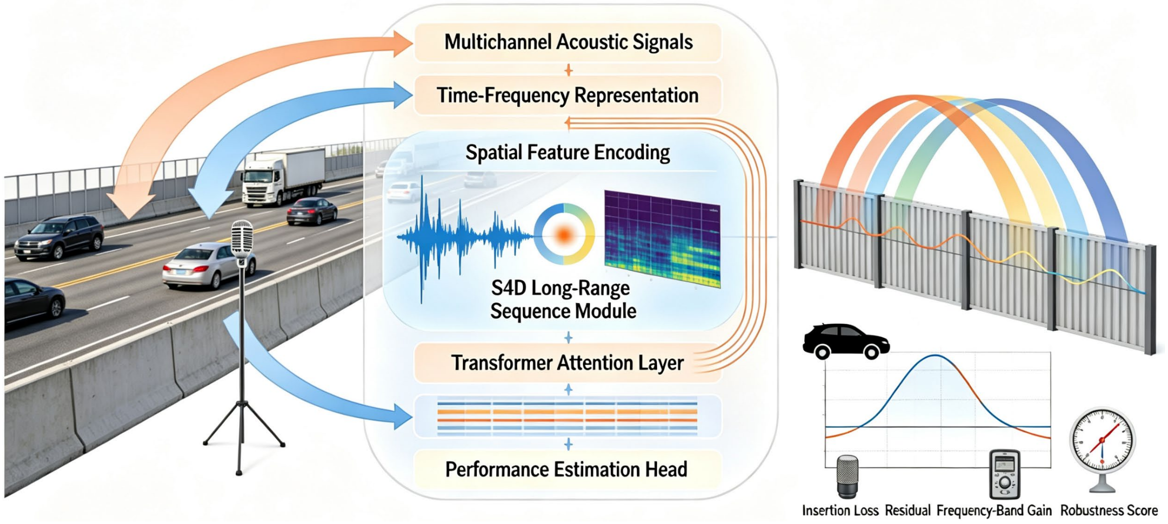

Figure 2. S4D-Transformer Architecture for Source-Receiver Acoustic Sequence Fusion.

After S4D encoding, the transformer fusion layer calculates multi-head attention through the following scaled dot-product operation:

$$A(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad \text{Eq.(4)}$$

Evaluation Objective and Implementation Details

A multi-objective loss function is employed during training to combine insertion-loss regression, octave-band attenuation regression and uncertainty calibration. Insertion loss term corrects the engineered rating; the spectral term keeps the diagnostic value; the uncertainty term avoids overconfidence in abnormal traffic or weather. The loss weights were chosen from a validation set and then fixed for all the reported experiments. No reference citations are used in this chapter because the formulation and implementation are presented as the proposed methods.

The multi-objective training loss is defined as:

$$L = \lambda_1 \|y - \hat{y}\|_1 + \lambda_2 \|s - \hat{s}\|_2^2 + \lambda_3 L_{cal} \quad \text{Eq.(5)}$$

It uses 96-dimensional feature tokens, four S4D blocks, four attention heads, and a two-layer feedforward prediction head in the implementation. Set the dropout rate to 0.10 and train the model using mini-batches of 32 windows for 120 epochs. Early stopping is performed after 15 consecutive epochs without an improvement in the validation loss. All acoustic features are standardized using statistics computed only from the training set. The predicted confidence interval is obtained from the learned variance head and compared with the reference measurement deviations.

The final window-level insertion-loss estimate and its variance are calculated as:

https://doi.org/10.64972/dea.2026.v5i3.3961d:1-14

$$\hat{I} = mean(\hat{y}_{1:T}), \sigma_I^2 = mean\left(exp(\hat{v}_{1:T})\right) \quad \text{Eq.(6)}$$

Experimental Data and Performance Analysis

Field dataset and experimental protocol

A field experiment was set up on a six-lane urban expressway adjacent to a 4.2-metre-high absorbent barrier. Two eight-channel arrays were installed 1.5m high above the ground, and one array was placed 3.0m away from the traffic-side barrier face, while the other was located 12.0m behind the receiver-side face. Over 28 days, the monitoring campaign produced a total of 1,344 ten-minute windows after quality inspection. The retained data includes 7.8 million valid array frames, 46,210 recorded vehicle passes detected acoustically, and 9,684 heavy vehicle events. Following the paired-measurement logic commonly used in barrier-performance assessment [24], reference insertion loss was obtained from calibrated source-side and receiver-side measurements of 126 manually inspected windows, and the remaining windows were used for model training and blind evaluation under the same preprocessing chain. The dataset was divided into 70% training, 10% validation and 20% test windows by day instead of randomly dividing frames, and adjacent traffic states were thus not leaked across the partitions.

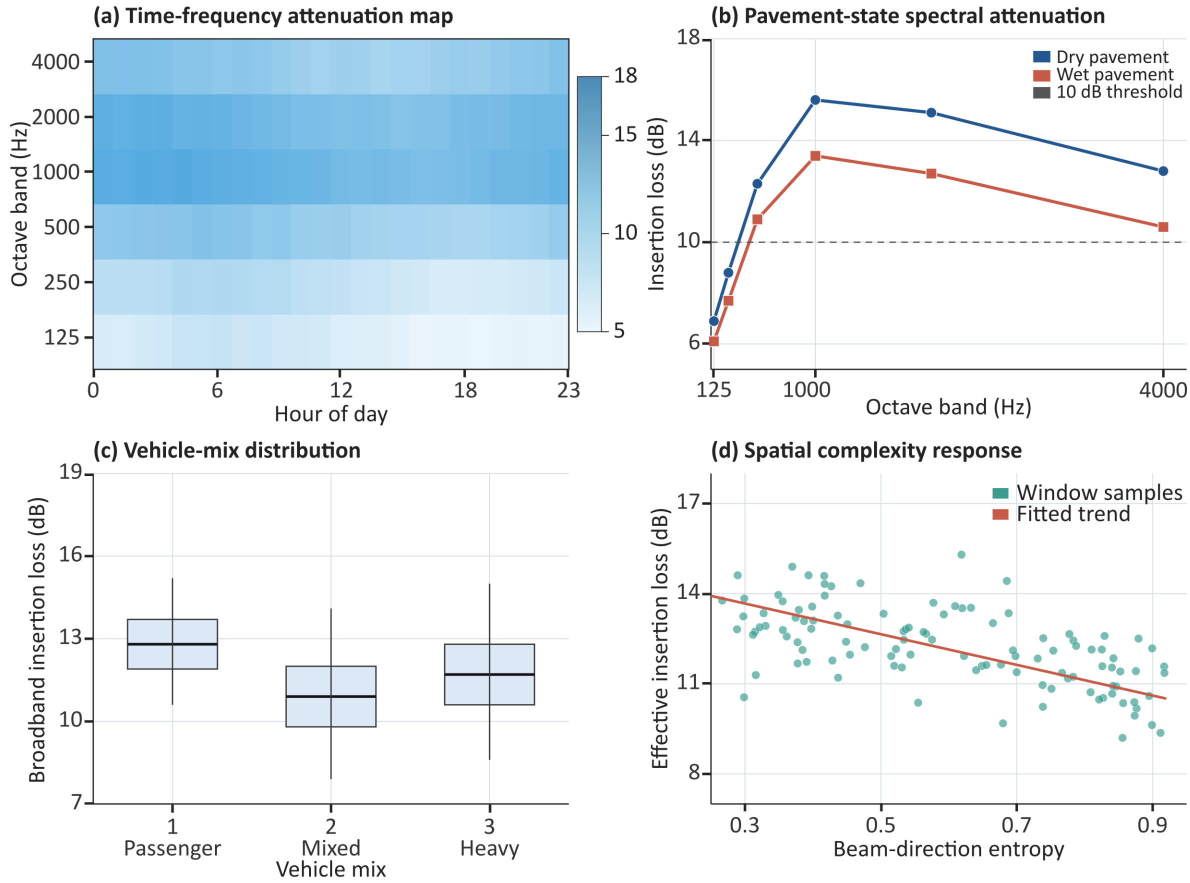

Figure 3. Spectral and spatial characteristics of measured barrier attenuation: (a) time-frequency attenuation map, (b) octave-band attenuation under pavement states, (c) insertion-loss distribution by vehicle mix, and (d) beam-direction entropy versus effective insertion loss.

A delay in the barrier's response occurred at a particular frequency. The median broadband insertion loss during the free-flow passenger-car interval was 12.8 dB, and under mixed-traffic conditions with more than 18 per cent heavy vehicles, it dropped to 10.9 dB. As low-frequency diffraction at the edge of a barrier is a well-known source of residual receiver-side energy [25], the 125 Hz and 250 Hz bands retained a relatively high level in the measured response. Attenuation increased at 1 kHz and 2 kHz when the receiver-side array was not directly in line-of-sight of propagation. Figure 3(a) is the time-frequency attenuation map; a relatively large reduction occurs in the 800 Hz - 2.5 kHz band under stable traffic conditions, as shown in Figure 3(b); the 1 kHz band shows a decrease from 15.6 dB in dry pavement to 13.4 dB in wet pavement; Figure 3(c) reports an interquartile spread of 2.1 dB for heavy-vehicle windows; and Figure 3(d) shows a negative correlation between beam-direction entropy and effective insertion loss when multiple source lanes are active. The four views show that a model trained only on broadband sound level may not detect some barrier-related mechanisms.

Table 1 shows the monitored acoustic and traffic variables in the model. The mean source-side equivalent level was 76.4 dB, and after calibration, the mean receiver-side was 63.9 dB. The average share of heavy-vehicles was 13.7%, and the day-level standard deviation was 5.4 percentage points. The mean wind speed was 2.1 m/s, and windows exceeding 4.0 m/s were only kept if the array coherence screening showed stable propagation. The above statistics are used to normalize the sequence features and to set the operating range for the reported evaluation. The model was not extended to the geometry of the observed corridor or the measured barrier height during training.

Table 1. Monitored variables and descriptive statistics of the roadside acoustic-array dataset

Variable	Unit	Mean	SD	Range	Model role
Source-side LAeq	dB	76.4	4.9	63.2-88.7	Incident field normalization
Receiver-side LAeq	dB	63.9	5.2	50.1-77.4	Shielded-field response
Broadband insertion loss	dB	12.5	2.6	7.3-18.9	Primary target
Heavy-vehicle share	%	13.7	5.4	2.1-28.6	Traffic-state descriptor
Array coherence	-	0.71	0.12	0.38-0.94	Spatial field stability
Wind speed	m/s	2.1	0.9	0.2-4.8	Propagation screening

The reference windows cover the morning peak hour, evening peak hour, off-peak free-flow hour and late-night heavy-vehicle passage hour. This distribution is relatively insensitive to changes in the acoustic environment at the same broadband insertion loss. For example, a 12 dB reduction in the speed of passenger-car free-flow generally showed stable shielding in the 500-Hz to 2-kHz band, but a similar broadband value obtained under truck-dominated traffic still contained substantial low-frequency residual energy. Manual inspection of the reference windows also found a small number of intervals affected by local reflections behind the receiver array. An interval was chosen to meet the general characteristics of a roadside area, and in a practical evaluation model, the effect of the barrier on noise generated by nearby hard surfaces had to be isolated.

Quality control discarded 4.6% of the original windows. The reasons for rejection were, in order of frequency: rain damage to the protective microphone grids, short communication interruptions, and abnormal coherence collapse caused by temporary road work. No window was removed solely because the insertion loss was small. Thus, it avoided the problem of exclusion due to poor performance in difficult cases from the dataset. The median synchronization error of the retained windows after clock correction was less than 0.7ms, and the drift in microphone sensitivity did not exceed 0.4dB during the campaign. These values are relatively small compared with the measured barrier attenuation range, but they were still propagated to the confidence target during training.

Meteorological factors served as screening and interpretation indices in the experimental protocol but were not the main predictors. Temperature, humidity and wind were sampled at 1-minute intervals, but the model input still only used acoustic-array data. Therefore, a deployable evaluation can be carried out at the site where a full meteorological station is not always available. Meteorological labels were used in the analysis to divide errors and to identify windows in which acoustic confidence should be interpreted cautiously. Separating the model input from the interpretation metadata also reduced the risk of learning a site-specific weather shortcut. A barrier should not be deemed effective simply because a certain weather condition is associated with lower traffic noise; it should be judged based on the measured contrast of acoustic fields at the incidence and reception ends.

Model comparison and ablation analysis

The S4D-Transformer was compared with support vector regression, random forest, CNN-LSTM, vanilla Transformer, pure S4D, and a transformer without source-side features. All neural baselines used the same train-validation-test partitions and the same calibrated feature inputs unless a feature was intentionally removed. The baseline set covers kernel regression, ensemble learning, recurrent spectral modeling, attention-based sequence modeling, and state-space sequence modeling, which are common comparison families in data-driven acoustic evaluation [26]. The proposed model achieved a broadband mean absolute error of 0.84 dB and an RMSE of 1.12 dB, compared with 1.35 dB and 1.78 dB for CNN-LSTM and 1.18 dB and 1.55 dB for the vanilla Transformer. The pure S4D model reached 0.97 dB MAE, indicating that long-memory sequence modeling was valuable even without attention-based fusion. Figure 4(a) compares model errors across broadband, low-band, mid-band, and high-band targets, where the proposed model reduces mid-band MAE to 0.79 dB against 1.15 dB for pure S4D and 1.27 dB for vanilla Transformer; Figure 4(b) shows the distribution of absolute window errors, with the proposed model keeping the 90th percentile below 1.92 dB. The gain was largest during heavy-vehicle windows because the attention layer assigned higher weight to channel-coherence changes and source-side spectral slope.

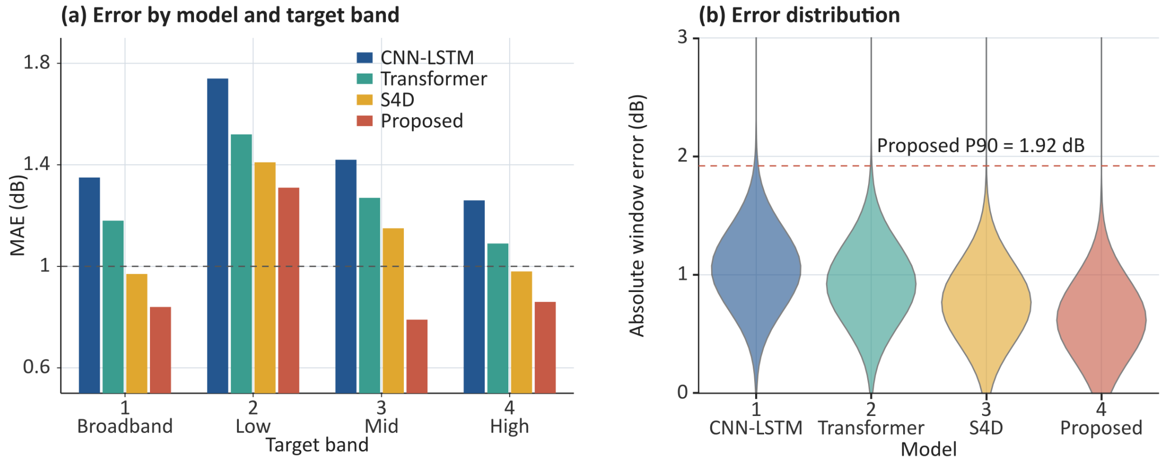

Figure 4. Baseline model comparison: (a) grouped error metrics by target band and (b) distribution of absolute window-level errors.

Ablation experiments determine the functions of parts by exclusion. Removal of the S4D layer increased the broadband MAE from 0.84 dB to 1.16 dB due to a lack of long-term traffic-state drift representation. Transformer fusion increased the error to 0.97 dB and reduced the high-band prediction for complex receiver-side reflections noticeably. Ablation analysis is often performed to determine the architectural and feature-level contributions of sequence models [27]. Eliminating spatial-coherence features increased the high-band MAE by 0.21 dB, and removing source-side incident features caused the largest degradation, raising the broadband MAE to 1.42 dB. Therefore, it cannot be solved by adding only receiver-side noise prediction. Table 2 shows the ablation results and parameter counts. The entire model has 2.8 million trainable parameters and requires 18.6 ms per ten-minute window on a laptop-grade GPU after feature extraction; it is suitable for inspection-scale monitoring.

Table 2. Ablation performance on the blind test partition

Model variant	MAE (dB)	RMSE (dB)	r	95% coverage (%)	Params (M)
Full S4D-Transformer	0.84	1.12	0.946	93.8	2.8
Without S4D memory	1.16	1.51	0.902	91.5	2.5
Without transformer fusion	0.97	1.31	0.928	92.1	2.1
Without spatial coherence	1.03	1.38	0.919	91.9	2.7
Without source-side features	1.42	1.88	0.861	88.4	2.4
Without uncertainty head	0.88	1.19	0.939	84.6	2.6

Training stability will also be addressed, and environmental acoustic data often contain outliers. The proposed loss reduced the validation oscillation after epoch 40 and avoided the overfitting of the vanilla Transformer. Remove the uncertainty head, and the point-estimate MAE rises slightly to 0.88 dB from 0.84 dB, but the empirical coverage of the 95% interval drops to 84.6% from 93.8%. Uncertainty-aware acoustic monitoring is applicable to model outputs that will be used for inspection or maintenance decisions [28]. In this experiment, such calibration was necessary because the maintenance decision might be based on whether an observed low insertion-loss interval is a real barrier defect or a high-uncertainty traffic condition. Therefore, the uncertainty output was kept as an inspection tool and not substituted for the predicted acoustic indicator.

Error decomposition also showed that the proposed architecture was less susceptible to traffic imbalance than the baselines. In windows with a heavy-vehicle share of less than 8%, the full model and the pure S4D model varied by only 0.08 dB in MAE, because the acoustic scene was dominated by stable passenger-car tire-road noise. At a proportion of heavy-vehicles of 18%, the difference was 0.24 dB. The transformer fusion layer was thus more effective in addressing the problems of low-frequency source energy, receiver-side coherence loss and beam entropy simultaneously. Occasionally, the vanilla Transformer produced accurate predictions for these windows, but its variance was relatively high; several high-error cases occurred when a single truck pass-by received excessive attention despite a normal source-receiver contrast in the surrounding window.

Feature extraction is used to determine the computational load, and a continuous acoustic front-end can run on a low-power chip. In a 10-minute window, the CNN-LSTM baseline had a latency of 11.3ms; the vanilla Transformer was 24.7ms; the pure S4D model was 14.9ms; and the proposed S4D-Transformer reached 18.6ms. Therefore, the added cost of attention was relatively small as it operated on compact state tokens instead of raw frames. The largest-memory-consumption point in inference is 412MB, which is still within the range of a workstation or a roadside inspection unit. The above time values show that the model can support near-real-time reporting for many barriers along a corridor that are monitored sequentially.

Error cases were reviewed to determine where the proposed model still performed poorly. The largest positive errors occurred when the receiver-side reflections temporarily increased the apparent shielded level and the source-side array recorded a stable incident field. In such cases, the model sometimes assumed that the reduction in barrier effect was due to a lower local receiver environment. The largest negative error occurs during sparse nighttime traffic periods, when only a few high-energy passes are within the window. S4D memory has improved the situation compared with the baseline, but the evidence is still less substantial than in dense traffic. These cases suggest that in the future, a small diagnostic tag can be stored with each estimate, including traffic density, coherence stability and reflection likelihood, so that engineers can determine whether a low score is due to a barrier itself or surrounding measurement conditions.

Frequency-resolved performance and engineering interpretation

Frequency-resolved analysis shows that the model learns physically plausible barrier behavior. Diffraction and ground interaction are known to complicate low-frequency barrier assessment [29]. Therefore, all the models showed a larger error in the low-frequency band below 250 Hz because these mechanisms reduced the contrast between source-side and receiver-side arrays. Despite this, S4D-Transformer's 125 Hz MAE was 1.31 dB, and it was better than CNN-LSTM's 1.74 dB. At 500 Hz, 1 kHz and 2 kHz, the errors of the proposed model were less than 0.90 dB, and the correlation coefficients exceeded 0.94. Figure 5(a) presents the predicted and measured octave-band insertion loss, where the proposed curve follows the measured peak near 1.6 kHz; Figure 5(b) shows that residual errors are centered near zero with a standard deviation of 0.98 dB; Figure 5(c) compares measured and predicted values with a regression slope of 0.97; and Figure 5(d) gives the cumulative error distribution, where 82.4% of test windows fall within 1.5 dB. The spectral behavior is consistent with the monitored barrier type, which combines absorptive panels with a continuous concrete base.

The model has also been run in operation. Wet and slippery road surfaces alter the tire-road spectral characteristics and high-frequency radiation properties [30]. The current measurements have added high-frequency content at the source and reduced the apparent high-band barrier benefit in the uncorrected data. Large crosswind fluctuations led to an extended uncertainty range, despite a relatively small mean-prediction error. Figure 6(a) compares dry, wet, low-wind, and crosswind conditions, showing that crosswind windows have a median predicted insertion loss of 11.2 dB and an interquartile range of 2.4 dB; Figure 6(b) maps the contribution of source level, receiver level, coherence, and traffic descriptors to the prediction, with receiver-side octave energy contributing 31% and source-receiver coherence contributing 22%. The stressor results suggest that model interpretation should be tied to measured acoustic evidence rather than treated as a black-box rating.

https://doi.org/10.64972/dea.2026.v5i3.3961d:1-14

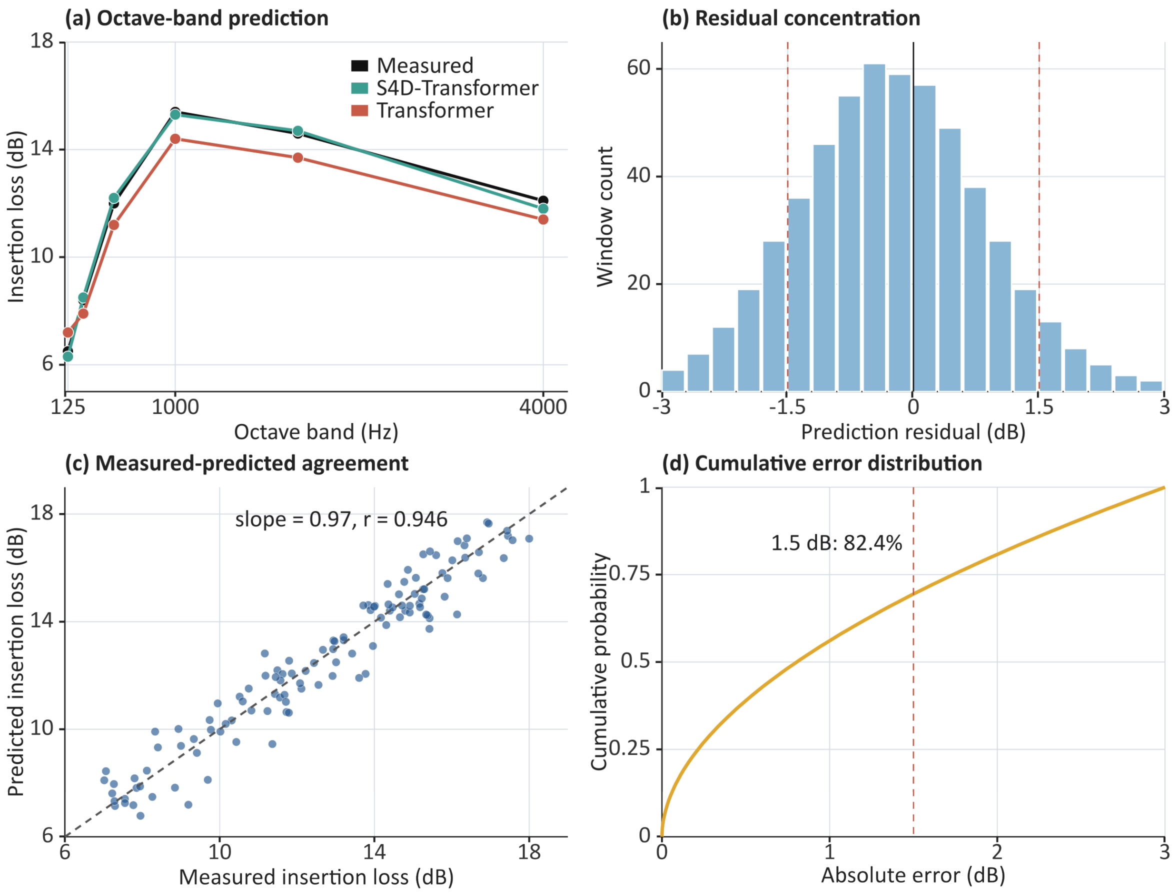

Figure 5. Frequency-resolved prediction behavior: (a) octave-band measured and predicted insertion loss, (b) residual-error distribution, (c) measured-predicted agreement, and (d) cumulative error distribution.

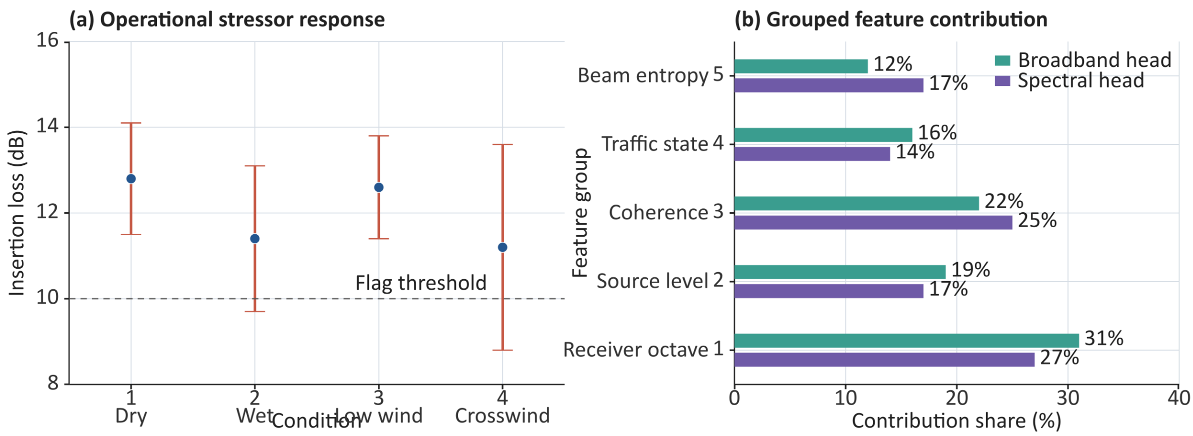

Figure 6. Operational stressor interpretation: (a) insertion-loss variation under pavement and wind states and (b) grouped feature contribution to broadband and spectral predictions.

Finally, the trained indicators helped to provide support for a maintenance-oriented decision. A window was triggered when the predicted broadband insertion loss fell below 10 dB and the confidence interval width was less than 2.5 dB. Maintenance-oriented acoustic screening generally needs to convert continuous measurements into reviewable event groups [31]. Under this rule, 9.7% of the test windows were flagged, and manual inspection showed that most of the flags occurred during periods of high low-frequency leakage or strong receiver-side reflections near a service opening. Figure 7(a) shows the day-by-day barrier performance index, where three clusters of low ratings appear after rainy evenings; Figure 7(b) separates flagged windows by dominant spectral cause, with low-frequency diffraction accounting for 41% and mid-band leakage for 34%;

https://doi.org/10.64972/dea.2026.v5i3.3961d:1-14

Figure 7(c) reports calibration reliability, where predicted uncertainty closely follows empirical error up to the 90th percentile; and Figure 7(d) places the proposed method in a maintenance matrix that combines insertion loss and confidence. Decision matrices are useful for separating acoustic severity from confidence before field inspection is scheduled [32]. The matrix did not automatically order the repairs, but it reduced the number of windows needing manual inspection in the test set from 269 to 51. This operating mode uses the general inspection logic of continuous screening before detailed on-site diagnosis [33].

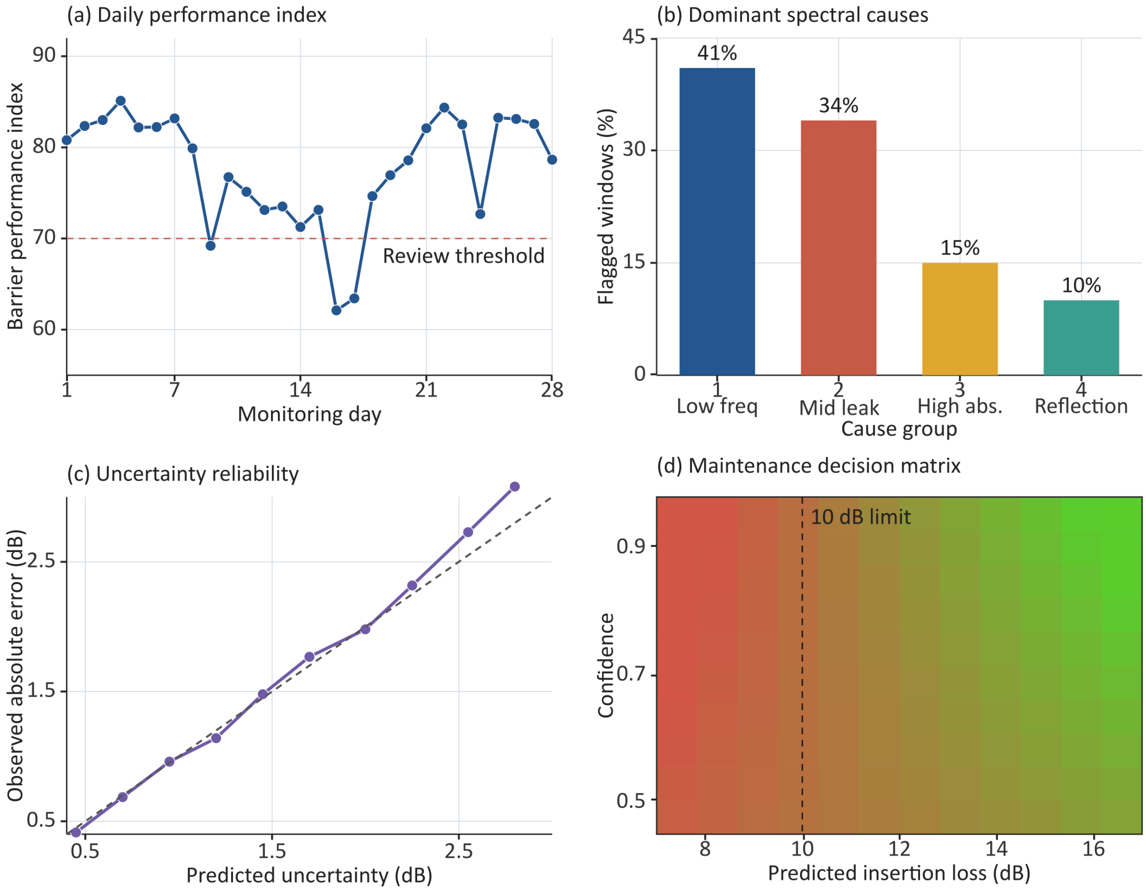

Figure 7. Maintenance-oriented evaluation outputs: (a) daily barrier performance index, (b) spectral causes of flagged windows, (c) uncertainty reliability, and (d) insertion-loss confidence maintenance matrix.

As shown in the enlarged windows above, it can be done to obtain real-time data for field work. Low-frequency diffraction flags were generally associated with high source-side energy below 250 Hz and only a small reduction in receiver-side coherence, suggesting that the energy bent over the barrier rather than passing through a local defect. The distribution of mid-band leakage flags was not uniform; receiver-side coherence was relatively high near 500 Hz and 1 kHz, and the beam-direction estimate was concentrated in a narrow angle. Such windows are suitable for observing the panel joints, service doors and drainage discontinuities. High-frequency absorption flags were less frequent, but they tended to occur after a wet period when the receiver-side spectrum above 2 kHz recovered faster than expected. The model does not directly identify a physical defect; it reduces the acoustic evidence to categories for inspection priority.

The confidence output is intended to support engineering judgement rather than replace field inspection. A small predicted insertion loss with a wide confidence interval was not considered a severe underperformance event because it often occurred in the presence of crosswinds or irregular traffic composition. Low insertion loss and a narrow uncertainty indicated that the source-receiver contrast, coherence pattern and spectral shape were all consistent with the same judgment. Thus, some of the unnecessary inspection alarms were eliminated, and only those likely to be long-term problems were retained. Therefore, the operating level of confidence output is not just a statistical supplement. It helps distinguish measurement uncertainty and acoustic evidence that is strong enough to support a follow-up.

The frequency-resolved outputs are also useful for comparing barrier segments over time. A broadband indicator may remain nearly unchanged while the spectral pattern shifts from high-frequency absorption loss to low-frequency diffraction dominance. Such a shift would imply a different inspection response. In the tested corridor, several low-index evenings did not show a uniform reduction across all bands; instead, the 500 Hz to 1 kHz range weakened while the 2 kHz band remained near its normal attenuation level. This pattern is compatible with localized leakage or geometric discontinuity rather than broad material degradation. By presenting the spectral estimate together with uncertainty, the model gives the engineer a compact description of both magnitude and likely acoustic mechanism.

Conclusion

This paper proposes a roadside acoustic-array data-driven S4D-Transformer model to assess the noise barrier performance under real-world traffic conditions. Combine synchronized source-side and receiver-side arrays, frequency-resolved acoustic descriptors, a diagonal state-space temporal memory, and transformer-based fusion to estimate insertion loss, spectral attenuation and confidence. Based on the field study, the model will have changed traffic patterns over an extended period and retain the spectral and spatial information required for engineering diagnosis. Thus, the constructed framework can be used repeatedly to conduct an extended evaluation of the barriers.

Several limitations remain. The dataset was collected at a single expressway corridor and one barrier type, so the model should not be considered universally applicable to all geometries, materials or climates. Extreme weather conditions were filtered out instead of being fully modelled, and the current feature set does not explicitly include visual traffic classification or detailed vehicle trajectories. A simple expression is provided by an uncertainty head; however, regular reference calibration must still be carried out because of sensor ageing, microphone replacement or alterations in the roadside geometry.

Future work will extend the framework to multiple barrier heights, absorbent materials and road alignments. A larger multi-site dataset can be employed to test the generalization capabilities of transfer learning and domain adaptation. Future deployments should add a small weather station and privacy-preserving traffic classification to better distinguish barrier degradation from changes in the source. Another is to integrate the output of the model with maintenance planning tools and, taking acoustic evidence, inspection costs and community disruption into account, adopt a single decision-making process.

Author Contributions

Sultan Al-Dhaheri contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. Khamis Al-Zahrani contributes to validation, analysis, investigation, data collection, draft preparation. Hazaa Al-Mansoori contributes to software, validation, analysis, investigation. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

Declaration of generative AI and AI-assisted technologies

During the preparation of this manuscript, the authors utilized Grammarly for linguistic polishing, logical organization and draft revision. The generative AI tool was not involved in research design, data collection, data analysis, or the equation of core conclusions. All outputs generated by AI were reviewed, revised and verified manually by the authors. The authors take full responsibility for all content of this manuscript.

References

[1] Rossi, D., Mascolo, A., Singh, D., & Guarnaccia, C. (2025). Machine Learning Regressors Calibrated on Computed Data for Road Traffic Noise Prediction. Machine Learning and Knowledge Extraction, 7(4), Article 133. https://doi.org/10.3390/make7040133

[2] Montenegro, A. L., Rey-Gozalo, G., Arenas, J. P., & Suárez, E. (2024). Streets classification models by urban features for road traffic noise estimation. Science of the Total Environment, 932, Article 173005. https://doi.org/10.1016/j.scitotenv.2024.173005

[3] Galiana Nieves, J., Picó, R., Del Rey, R., Alba, J., & Redondo, J. (2024). Assessment of the sound reduction index provided by noise barriers with low sound insulation. Applied Acoustics, 220, Article 109967. https://doi.org/10.1016/j.apacoust.2024.109967

[4] Li, F., Lai, R., Rong, Y., Yu, F., Du, C., Lan, Z., Ye, B., & Li, Z. (2024). Dynamic traffic noise simulation at signal-controlled intersections based on cellular automata model. Applied Acoustics, 224, Article 110128. https://doi.org/10.1016/j.apacoust.2024.110128

[5] Ramírez-Solana, D., Galiana-Nieves, J., Picó, R., Redondo, J., Sangiorgio, V., Vito Graziano, A., & Parisi, N. (2024). Sonic Crystal Noise Barrier with Resonant Cavities for Train Brake Noise Mitigation. Applied Sciences, 14(7), Article 2753. https://doi.org/10.3390/app14072753

[6] Leach, F. C. P., Yang, J., Shankar, V., Cole, S., Parnell, K., Entezami, M., & Bartington, S. E. (2024). The impact of a low traffic neighbourhood intervention on urban noise measured with low-cost sensors in Oxford, UK. Transportation Research Part D: Transport and Environment, 134, Article 104357. https://doi.org/10.1016/j.trd.2024.104357

[7] Acosta, Ó., Montenegro, C., & González Crespo, R. (2024). Road traffic noise prediction model based on artificial neural networks. Heliyon, 10(17), Article e36484. https://doi.org/10.1016/j.heliyon.2024.e36484

[8] Alam, P., Ahmad, T., Maqbool, L., Islam, R., Ahmad, S., & Sharholy, M. (2024). Noise pollution mitigation and control in urban areas near international borders through 2D noise mapping. Scientific Reports, 14(1), Article 30582. https://doi.org/10.1038/s41598-024-82940-4

[9] Deng, L., Rising, H. H., Gu, C., & Bimal, A. (2024). Effects of labyrinth-like path designs on mitigating stress response to traffic noise. Building and Environment, 265, Article 111914. https://doi.org/10.1016/j.buildenv.2024.111914

[10] Deswal, S., Bhardwaj, P., & Pal, M. (2025). Traffic noise modelling on urban arterial roads using mathematical and machine-learning techniques. Noise & Vibration Worldwide, 56(4-5), 207-215. https://doi.org/10.1177/09574565251333063

[11] Wang, Y., Lin, C., Gong, B., & Liu, H. (2025). Traffic noise evaluation at intersection using roadside light detection and ranging sensor. Transportation Research Part D: Transport and Environment, 144, Article 104750. https://doi.org/10.1016/j.trd.2025.104750

[12] Staab, J., Weigand, M., Schady, A., Droin, A., Cea, D., Dallavalle, M., Nikolaou, N., Valizadeh, M., Wolf, K., Wurm, M., Lakes, T., & Taubenböck, H. (2025). National road traffic noise estimation with ensemble learning and multimodal geodata. Transportation Research Part D: Transport and Environment, 149, Article 105063. https://doi.org/10.1016/j.trd.2025.105063

[13] Sallam, O., El-Rayes, K., Ignacio, E. J., Hajj, R., & Al-Ghzawi, M. (2025). Mitigating external noise of transverse rumble strips and complaints from nearby neighborhoods. Transportation Research Part D: Transport and Environment, 149, Article 105021. https://doi.org/10.1016/j.trd.2025.105021

[14] Pantazatou, K., Mattisson, K., Olsson, P.-O., Telldén, E., Kettisen, A., Hosseinvash Azari, S., Liu, W., & Harrie, L. (2025). Influence of land cover on noise simulation output - A case study in Malmö, Sweden. Noise Mapping, 12(1), Article 20250016. https://doi.org/10.1515/noise-2025-0016

[15] Park, S. H., Lee, P. J., Jeong, J., & Kim, Y. (2025). Railway and road traffic noise annoyance of residents in high-rise apartments. Transportation Research Part D: Transport and Environment, 146, Article 104856. https://doi.org/10.1016/j.trd.2025.104856

[16] Jovanoska-Mitrevska, M., Samardzioska, T., & Mickovski, S. B. (2026). Application and optimization of reinforced concrete noise barrier. Buildings, 16(4), Article 745. https://doi.org/10.3390/buildings16040745

[17] Zhao, Y., He, Y., Chen, H., Zhang, Z., & Xu, Z. (2025). Three-dimensional grid-free sound source localization method based on deep learning. Applied Acoustics, 227, Article 110261. https://doi.org/10.1016/j.apacoust.2024.110261

[18] Müller, K., Buck, M., Doclo, S., Østergaard, J., & Wolff, T. (2025). A steered response power method for sound source localization with generic acoustic models. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 33, 4004-4019. https://doi.org/10.1109/TASLPRO.2025.3611789

[19] Pi, R., & Yu, X. (2025). Uncertainty estimation for sound source localization with deep learning. IEEE Transactions on Instrumentation and Measurement, 74, Article 2502512. https://doi.org/10.1109/TIM.2024.3522632

[20] Zhang, G., Geng, L., Xie, F., & He, C.-D. (2024). A dynamic convolution-transformer neural network for multiple sound source localization based on functional beamforming. Mechanical Systems and Signal Processing, 211, Article 111272. https://doi.org/10.1016/j.ymssp.2024.111272

[21] Wang, Y., Deng, Z., Zhao, J., Kopiev, V. F., Gao, D., & Chen, W.-L. (2025). Progress in beamforming acoustic imaging based on phased microphone arrays: Algorithms and applications. Measurement, 242, Article 116100. https://doi.org/10.1016/j.measurement.2024.116100

[22] Feng, D., Yu, L., Wei, L., Shi, Y., Pan, W., & Li, M. (2024). Acoustic inversion method based on the shear flow Green's function for sound source localization in open-jet wind tunnels. Mechanical Systems and Signal Processing, 220, Article 111650. https://doi.org/10.1016/j.ymssp.2024.111650

[23] Wu, X., Ni, R., Cui, X., Guo, T., & Peng, H. (2025). Sound source localization based on three-element microphone array and RepMobileViT model. Applied Acoustics, 231, Article 110480. https://doi.org/10.1016/j.apacoust.2024.110480

[24] Gao, L., Mao, Q., & Dong, M. (2024). On local temporal embedding for semi-supervised sound event detection. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 32, 1687-1698. https://doi.org/10.1109/TASLP.2024.3369529

[25] Guan, Y., Han, J., Song, H., Deng, S., Zheng, G., Zheng, T., & He, Y. (2024). Sound activity-aware based cross-task collaborative training for semi-supervised sound event detection. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 32, 4479-4492. https://doi.org/10.1109/TASLP.2024.3451983

[26] Wu, Q., Wang, M., Hu, Y., Ma, X., & Ou, Z. (2026). A dual consistency training strategy for polyphonic sound event detection. Neurocomputing, 673, Article 132880. https://doi.org/10.1016/j.neucom.2026.132880

[27] Barahona-Ríos, A., Naranjo-Alcazar, J., Perez-Castanos, S., Zuccarello, P., Cobos, M., & Ferri, F. J. (2024). Conformer-based sound event detection with semi-supervised learning and data augmentation. IEEE Access, 12, 13590-13602. https://doi.org/10.1109/ACCESS.2024.3355379

[28] Feng, C., Li, Z., & Wang, X. (2024). ElasticAST: An audio spectrogram transformer for all length and resolutions. Proceedings of the AAAI Conference on Artificial Intelligence, 38(16), 17734-17742. https://doi.org/10.1609/aaai.v38i16.29702

[29] Atito, S., Awais, M., & Kittler, J. (2024). ASiT: Audio spectrogram vision transformer for general audio representation. IEEE Signal Processing Letters, 31, 1224-1228. https://doi.org/10.1109/LSP.2024.3377915

[30] Cheng, L., Lei, D., & Tao, S. (2024). Recent advances in deep learning for traffic probabilistic prediction. Transport Reviews, 44(6), 1129-1135. https://doi.org/10.1080/01441647.2024.2408840

[31] Yu, Y., Ma, R., & Ma, Z. (2024). Robformer: A robust decomposition transformer for long-term time series forecasting. Pattern Recognition, 153, Article 110552. https://doi.org/10.1016/j.patcog.2024.110552

[32] Koya, S. R., & Roy, T. (2024). Temporal Fusion Transformers for streamflow prediction: Value of combining attention with recurrence. Journal of Hydrology, 637, Article 131301. https://doi.org/10.1016/j.jhydrol.2024.131301

[33] Zheng, P., & Zhi, Y. (2025). Design of differential microphone array beampatterns with sidelobe level constraints. Signal Processing, 227, Article 109748. https://doi.org/10.1016/j.sigpro.2024.109748