Mechanical Systems and Signal Processing 229 (2025) 112491

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

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

Echoformer: An echo state-embedded transformer for robust reconstruction of railway trackside noise on urban metro lines

Xin Ye[a,b,c,1], Yan-Ke Tan[d,e,f,1], Yi-Qing Ni[d,e,*]

[a] College of Civil Engineering and Architecture, Wenzhou University, Wenzhou 325035, PR China
[b] Key Laboratory of Engineering and Technology for Soft Soil Foundation and Tideland Reclamation of Zhejiang Province, Wenzhou 325035, PR China
[c] Zhejiang Engineering Research Center of Disaster Prevention and Mitigation for Coastal Soft Soil Foundation, Wenzhou, Zhejiang 325035, PR China
[d] Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom 999077, Kowloon, Hong Kong, PR China
[e] National Rail Transit Electrification and Automation Engineering Technology Research Center (Hong Kong Branch), Hung Hom 999077, Kowloon, Hong Kong, PR China
[f] College of Civil Engineering, Tongji University, Yangpu District, Shanghai 200092, PR China

ARTICLE INFO

Keywords:
Rolling noise
Signal reconstruction
Time series mapping
Echo state network
Transformer

ABSTRACT

Railway rolling noise on straight railway lines has become a crucial environmental impact of railway systems. The vibration of the rail tracks is the primary contributor to the formation of rolling noise. Developing a surrogate model to capture the reflectional relationship between track vibrations and trackside noise is desired in two perspectives. Firstly, it offers a solution for noise monitoring with data loss, or when field conditions are restrictive for sensors' deployment. Secondly, such a model can facilitate the design and optimization of noise control devices in laboratory, where the actual trackside noise is intricate to simulate. However, it is a dauting task to reveal the underlying relationship between track vibration and trackside noise. This work introduces Echoformer, a novel framework that blends echo states with the transformer architecture, designed to perform time series mapping. Comprehensive testing shows that the Echoformer outperforms conventional RNN architectures in reconstructing both near-field and far-field trackside noise. Moreover, the Echoformer exhibits remarkable resilience against information loss and noisy signal scenario, ensuring a robust reconstruction for the task. This study underscores the Echoformer's potential as a steadfast tool in the realm of railway noise analysis.

1. Introduction

Rolling noise on straight railway lines is one of the dominant factors that influencing the environment and residents in the vicinity of urban railway systems [1,2]. After a period of operation, rail corrugation can develop on the wheel-rail contact surface [3]. Consequently, vibrations originating from the irregular contact between the wheel and rail can induce rolling noise, leading to noise pollution issues. In the case of rolling noise, track vibrations are the primary source. The vibration spectrum mainly falls within the frequency range of 500 Hz to 1000 Hz [4]. At present, the rolling noise has become a key obstacle for the development of urban railway transportation [5]. Therefore, it is essential to monitor trackside noise in noise-sensitive areas, such as schools, hospitals, and

[*] Corresponding author.
E-mail address: ceyqni@polyu.edu.hk (Y.-Q. Ni).
[1] Co-first authors with equally contribution to this study.

https://doi.org/10.1016/j.ymssp.2025.112491
Received 28 September 2024; Received in revised form 22 January 2025; Accepted 18 February 2025
Available online 26 February 2025

residential neighborhoods. According to the measurement standard for railway noise (BS EN 15461:2008 + A1:2010), microphones are suggested to be set on both near field (0.6 m from the rail top) and far field (7.5 m away from the track centerline, and 1.2 m above the rail top) [7]. However, restricted by the safety operation regulations, the installation of near-field microphones must be prudently inspected [8]. For an urban rail viaduct section, an additional temporary structure is sometimes required to access the far-field measurement location. [9]. Such in-situ tests must be conducted with the landowner's permission. In conclusion, obtaining complete monitoring data is not always feasible. In such cases, a method to infer missing data using available sensors becomes an appealing approach for railway noise analysis.

Data loss is another challenge in noise monitoring projects. Occasional sensor malfunctions or failures, often caused by power outages or poor wiring connections, can result in the loss of the monitoring data [10–13]. When wireless sensors are used, data loss may occur during the transmission from the sensors to the receivers [14]. If a period of data is missing due to the aforementioned reasons, reconstructing the missing data is essential to maintain intact monitoring information.

On the other hand, controlling noise at its source, specifically rail track vibrations, is the most fundamental strategy for mitigating rolling noise. To this end, various rail track vibration control devices have been developed, including rail dampers [15–19], rail pads [20,21] and rail clips [22,23]. Optimizing these devices requires extensive tuning and validation tests. Clearly, constant in-situ trial and error is impractical. However, in the laboratory, researchers can only simulate rail track vibrations, as the noise is influenced by numerous field conditions that cannot be fully replicated in a laboratory environment. Therefore, a surrogate model that can predict trackside noise from rail track vibrations would greatly facilitate the development of noise and vibration control devices.

Reconstructing railway rolling noise is a dauting task [24]. There has been a recent surge in the application of deep learning (DL) methods owing to their rigorous ability to dissect nonlinear systems [25–32]. Translating track vibration signals into trackside noise is a typical sequence-to-sequence task. The recurrent neural network (RNN) is a potent tool that integrates large dynamic memory and adaptable computational abilities to recognize sequential information [33]. Embedded with delay loops in their neurons, the topological structure of RNNs allows them to generate a stacked flow for processing sequential data. In the development of RNNs, gating mechanism is applied in the gated recursive neuron to improve the long-term memory of the structure [34]. Thanks to this technique, long short-term memory (LSTM) network [35–37] and gated recurrent unit (GRU) [38,39] can link states over long distances. However, these RNNs are also inherently tedious to train due to their expanded parameter space. In light of this downside, a variant of RNN named echo state network (ESN) has garnered attention from the research community [40–42].

In the standard ESN architecture, input sequences are projected into a reservoir defined by a large-scale latent state space. Akin to the kernel method, this sparsely connected, high-dimensional state space can generate linearly separable features. Therefore, the training of an ESN can be simplified to a simple linear regression problem as claimed by Jaeger et al. [43,44]. This architecture demonstrates exceptional performance in reconstructing time series data. [45].

The transformer architecture has also been widely adopted in recent research on time series analysis [46–48]. The self-attention mechanism within the encoder and decoder blocks effectively emphasizes the sections that require greater focus [49]. Some studies report that combining different model structures is an effective approach to fuse merits across models [50]. For instance, integrating Transformers with conventional neural networks is anticipated to enhance performance by leveraging the strengths of each component [51–54]. Liu et al. introduced a bidirectional GRU embedding as a preprocessing step for the input context before feeding it into the Transformer encoder. In their machine reading comprehension task, the innovative RNN embedding effectively reduced the impact of incorrect word segments, setting a new state-of-the-art benchmark [53]. In a similar vein, Xia et al. created a framework conflating LTSM and Transformer for their task. Through comparisons, they concluded that the combination of RNN and Transformer outperforms both stacked hierarchical RNN-RNN and Transformer-Transformer networks [54].

Building on this line of research, this paper attempts to reconstruct trackside noise signals from recorded rail track vibrations. In the proposed framework, the input data is first projected into a sparsely connected, large-scale reservoir (echo state) to extract high-dimensional and sequential features. Next, the self-attention block within the Transformer structure captures correlations across different time intervals. This novel method, termed Echoformer, is then compared with baseline models to demonstrate its superiority in predicting trackside noise. Finally, the generalization ability of the trained model is validated using data from a distinct metro line monitoring project.

In summary, the main contributions of this paper are summarized as follows:

(1) A novel RNN-Transformer framework, called Echoformer, is proposed to achieve the intricate time series mapping from track vibrations to trackside noise. This method could facilitate trackside noise monitoring and the development of noise control devices.
(2) The reconstruction capability of Echoformer is elucidated by comparison with baseline methods in various scenarios. Besides, the robust Echoformer withstands situations when input information is incomplete, and when the input signal is contaminated by noise.
(3) A dataset stemmed from another urban metro monitoring project is adopted to verify the generalization ability of the Echoformer. The results demonstrate that the proposed method can reliably capture noise features across different metro lines.

The rest of this paper is organized as follows: Section 2 introduces the methods of the involved blocks, including ESN, self-attention, and the proposed Echoformer structure. Section 3 describes the details of the in-situ experimental project, which investigated a viaduct section of an urban metro line. Section 4 presents the reconstruction results by Echoformer and the baseline methods. Testing on the resilience of the reconstruction is also presented in this section. Finally, Section 5 offers the conclusions and suggests directions for future work Fig. 1.

2. Methodology of Echoformer

2.1. Echo state network

The echo-state network (ESN) has been extensively studied for its remarkable ability to learn complex nonlinear dynamics in sequential signals. The nonlinearity of the ESN is embedded in its activation function for the state updating. The reservoir, or the latent state space, is randomly generated at the initial time state. A standard ESN can be regarded as a three-layer structure, including the input layer, the reservoir, and the output layer (Fig. 2). The evolution of the latent state along the time axis is defined as:

$$\boldsymbol{s}_{t+1} = \lambda \boldsymbol{s}_t + (1-\lambda)\Psi_a(\boldsymbol{s}_t, \boldsymbol{x}_t, \boldsymbol{y}_t), \tag{1}$$

where $\boldsymbol{s}_t \in \mathbb{R}^{N_s}$ represents the latent state, $\boldsymbol{x}_t \in \mathbb{R}^{N_x}$ is the input series signal to the ESN, the relaxation coefficient $\lambda \in [0,1)$ aims to call back the memory from the last timestamp to pursue the continuity of the state evolution. $\boldsymbol{y}_t \in \mathbb{R}^{N_y}$ is the target signal at the current state. The activation function $\Psi_a(\bullet)$ provides nonlinear property to ESN. The subscript t presents the general time stamp.

$$\Psi_a(\boldsymbol{s}_t, \boldsymbol{x}_t, \boldsymbol{y}_t) = \tanh(\boldsymbol{W_{res}} \bullet \boldsymbol{s}_t + \boldsymbol{W_{in}} \bullet \boldsymbol{x}_t + \boldsymbol{W_{back}} \bullet \boldsymbol{y}_t), \tag{2}$$

aside from calling back the last state, the activation (take hyperbolic tangent function as an example) of the state space shown in Eq. (2) contains the information of all three parts. Here, $\boldsymbol{W_{res}} \in \mathbb{R}^{N_s \times N_s}$, $\boldsymbol{W_{in}} \in \mathbb{R}^{N_s \times N_x}$, and $\boldsymbol{W_{back}} \in \mathbb{R}^{N_s \times N_y}$ are independent weight matrices that are generated by the uniform distribution. The structure of ESN presented by Eqs. (1) and (2) is very similar to a standard RNN. Whilst in RNN, the weight matrices are updated during the training process. But the weights mentioned in Eq. (2) will not be updated once they are generated. The reservoir weights, $\boldsymbol{W_{res}}$, define the connectivity inside the latent space. As aforementioned, the $\boldsymbol{W_{res}}$ should be a sparse matrix to ensure a richer internal dynamic of the reservoir [43].

After the initialization of the reservoir, it should go through a certain free-running period before it is ready to generate the output signal. In this study, the free-running length is set as 100 timestamps.

$$\widehat{\boldsymbol{y}}_t = \boldsymbol{W_{out}} \bullet [\boldsymbol{x}_t, \boldsymbol{s}_t, \widehat{\boldsymbol{y}}_{t-1}] + \boldsymbol{b_Y}. \tag{3}$$

Eq. (3) presents the output layer of the ESN. Following the observation given by Lukoševičius and Jaeger [55], taking the input, current state, and the previous output together can improve the model accuracy. The model updating of the ESN is only focused on the $\boldsymbol{W_{out}} \in \mathbb{R}^{N_y \times (N_x + N_s + N_y)}$. The task is to find a linear map of the target signal from the randomly generated reservoir. Therefore, the training is relatively straightforward since the reservoir has extracted plenty of information. Considering a training series $\boldsymbol{Y} = (\boldsymbol{y}_1, \cdots, \boldsymbol{y_N}) \in \mathbb{R}^{N_y \times N}$ with N segments, where N is the length of the training series. This linear map can be determined by following the regularized optimization problem, which can be presented by:

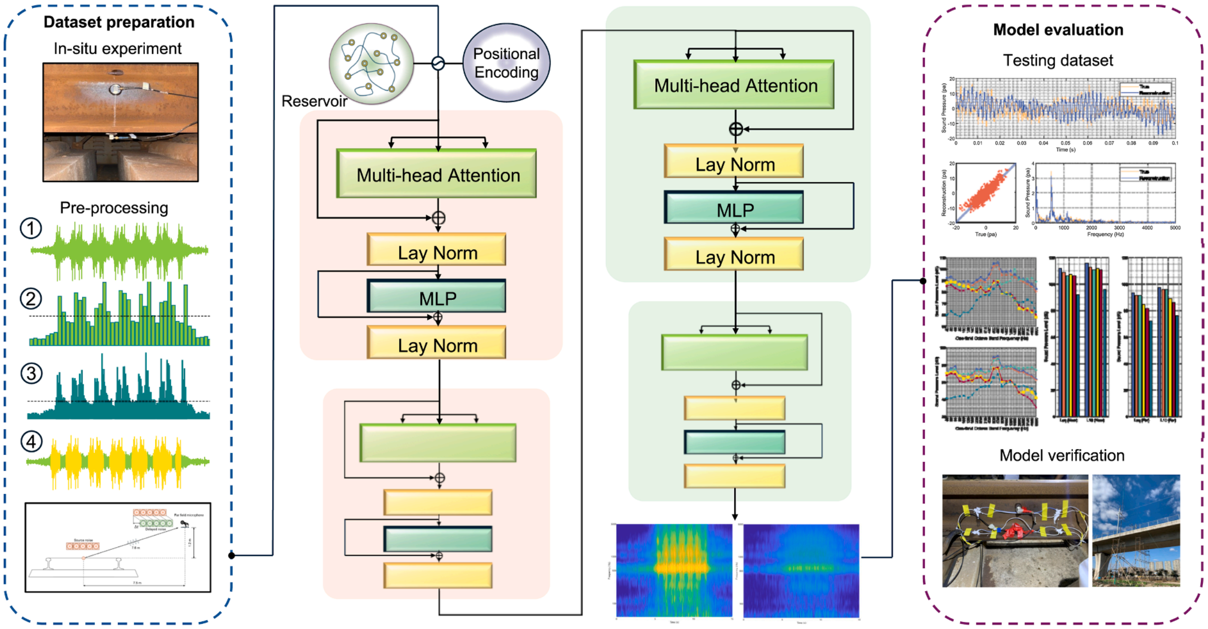

Fig. 1. Overall architecture of the proposed Echoformer.

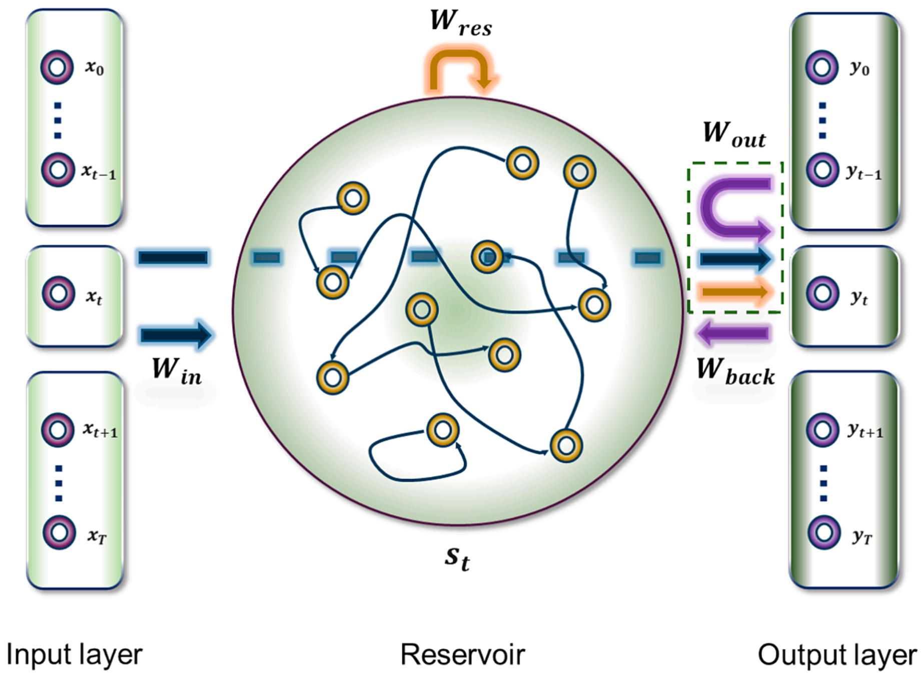

Fig. 2. Three-layer structure of the ESN.

$$\min_{W_{out}} \sum_{i=1}^{N} \frac{1}{2} \| \boldsymbol{y}_i - \widehat{\boldsymbol{y}}_i \|_2^2 + \frac{\beta}{2} \| \boldsymbol{W}_{out} \|_F^2. \tag{4}$$

In Eq. (4), $\| \bullet \|_F$ is the Frobenius norm, and the regularization parameter β ensures the sparsity of the calculated $\boldsymbol{W}_{out}$. The coefficient $1/2$ is adopted here to comfort the gradient computation.

2.2. Self-attention mechanism

To mimic the human attention ability, the self-attention mechanism has been proposed to distill valuable information from a mass of unimportant data [49]. Upon receiving the input series, the self-attention block (Fig. 3) assesses the importance of each component throughout the entire series.

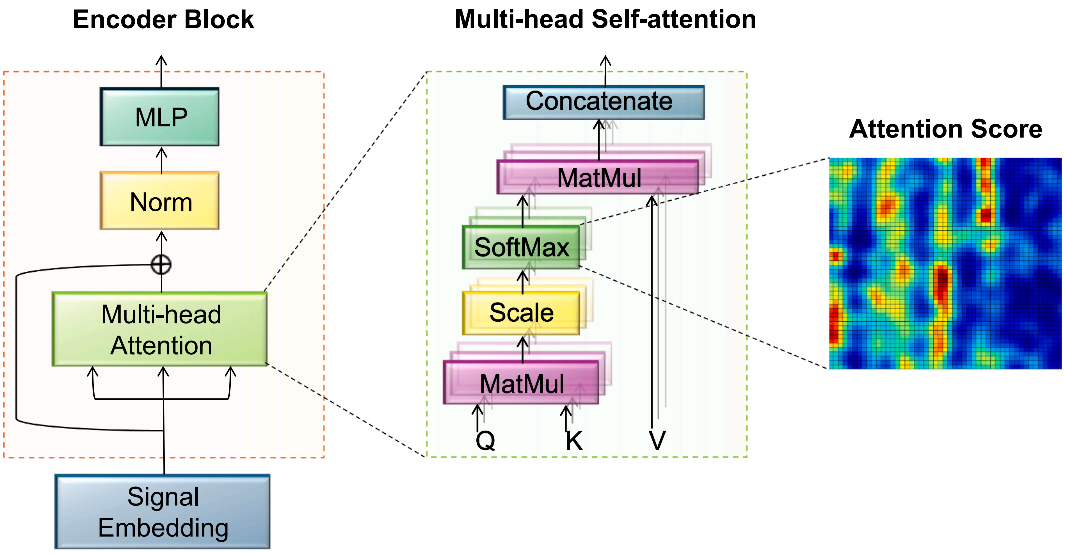

Fig. 3. Self-attention block in the Transformer structure.

Considering the input series $\boldsymbol{X} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N) \in \mathbb{R}^{N_x \times N}$ with length of N, the self-attention is achieved by three weight matrices: $\boldsymbol{W_Q} \in \mathbb{R}^{N_K \times N_x}$, $\boldsymbol{W_K} \in \mathbb{R}^{N_K \times N_x}$, and $\boldsymbol{W_V} \in \mathbb{R}^{N_V \times N_x}$, which represent the query, key and value weights, respectively. The attention score is calculated through a normalized SoftMax function, then applied to the weighted value matrix:

$$attention = softmax\left(\frac{(\boldsymbol{W_Q} \bullet \boldsymbol{X})^T \bullet \boldsymbol{W_K} \bullet \boldsymbol{X}}{\sqrt{N_K}}\right)(\boldsymbol{W_V} \bullet \boldsymbol{X})^T, \tag{5}$$

where the purpose of the normalization with $\sqrt{N_K}$ is to stabilize the gradient descend. Specifically, it can reduce the variance value of the SoftMax function's output. In Eq. (5), weight matrices $\boldsymbol{W_Q}$, $\boldsymbol{W_K}$, and $\boldsymbol{W_V}$ are trainable weights. Based on self-attention, multi-head self-attention applied multiple sets of queries, keys, and values that parallelly calculate the attention results. These multiple heads are eventually concatenated and then fed forward.

2.3. Framework of the Echoformer

Fig. 4 shows the flowchart of the proposed Echoformer. This structure comprises an encoder section and a decoder section. Each section could contain several encoder/decoder blocks. Similar to the Transformer structure, the encoder section receives the input series, and the decoder section decodes the prediction from information given by the encoder.

The time series data are first projected into the reservoir before being passed into the encoder section. This embedded reservoir is the same as the latent space in the ESN, which accumulates the high-dimensional sequence of the echo states. In our work, the size of the reservoir N_s is 1000. Positional encoding is integrated before the encoder to mark the relative location of all components within the series data [49].

Leveraging the power of multi-head attention, encoder blocks in the encoder section can process the entire input time series. Inside

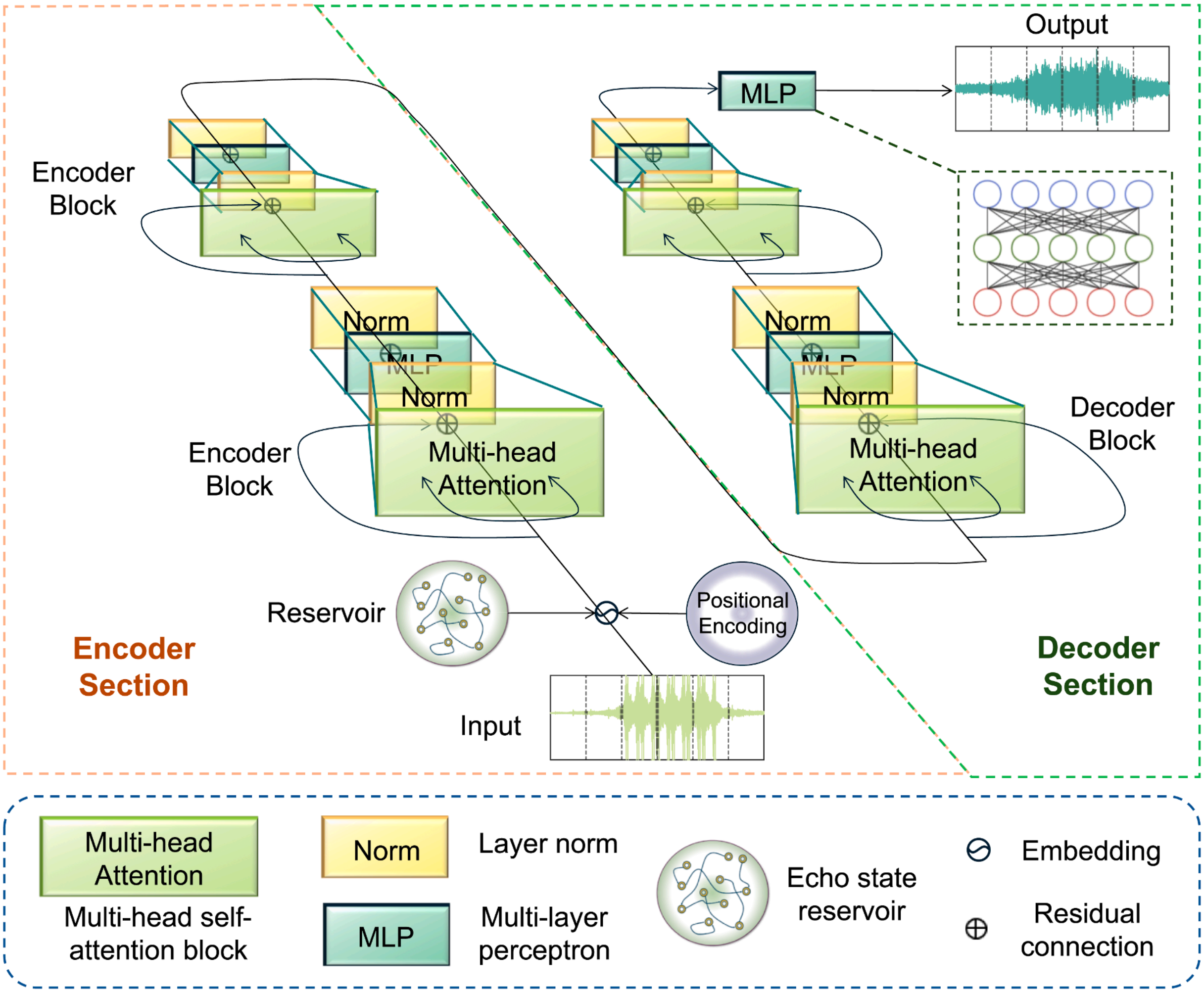

Fig. 4. Flowchart of the Echoformer.

encoder/decoder blocks, following the multi-head attention calculation, the results are fed forward through a multi-layer perceptron (MLP) in each block. Layer Normalizations [56] are adopted on all layers, and residual connections are implemented inside each block.

3. In-situ experiment on an urban viaduct metro line

3.1. Site environment and experimental setup

The in-situ experiment aimed to monitor the trackside noise on an urban viaduct metro line. The test section was between Tanglang station and University town station on the Shenzhen metro line 5. Due to the proximity of the residential area (Fig. 5(a)), this location is considered a noise-sensitive zone. The operation trains on this line are six-cabin trains (urban metro train type A, according to the Chinese railway standard TB-10624), and it passes the testing section at a speed of around 70 km/h. Since this is a straight metro line section, the predominant noise type is the rail rolling noise induced from the wheel-rail contact.

Triaxial accelerometers (Dytran, 3023A4), labeled A1, A2, A3 and A4, were attached to the rail tracks to measure the vibration responses. The acceleration was measured at rail web and rail bottom of both tracks. The deployment of accelerometers is presented in Fig. 5(b). Although these sensors recorded responses in three directions (vertical, lateral and along rail tracks), only the vibrations in vertical and lateral directions will be considered in this experiment.

Microphones (B&K, Type 4189) were adopted to monitor the trackside rolling noise. The first microphone (M1) was placed the same height as the rail top and 0.6 m away to record the near field noise. For the far field noise, the location of the microphone (M2) was 1.2 m above the rail top, and 7.5 m away from the center line of rail tracks. To facilitate the mounting of M2, a scaffold was set up near the viaduct to support a lightweight pole, to which the microphone could attach. In this in-situ experiment, the sampling rate of all sensors was set as 10000 Hz, since the dominant frequency of rolling noise is under 2000 Hz.

3.2. Experimental results on trackside noise and track vibration

The trackside noise is the target to be reconstructed in this work (Fig. 6). The most significant noise appears at the frequency around 630 Hz. The peak sound pressure level is 111.0 dB(A) at the near field and 93.8 dB(A) at the far field. The overall sound pressure levels at near and far fields are 95.8 dB(A) and 78.4 dB(A), respectively.

There are clear spikes on the recorded sound pressure when wheels pass through the testing section. According to the condition when wheels are passing or not, the time windows are divided into two categories, named as “with wheels” and “without wheels”.

Four accelerometers, namely A1, A2, A3 and A4, record acceleration in vertical (V) and lateral (L) directions. Therefore, there are eight channels of acceleration signal (Fig. 7). Rail track vibration energy is concentrated in the vertical direction according to the time–frequency spectrum of acceleration signals. While there are minor differences in the vibration of different tracks, sensor signals from different locations (web and bottom) on one track are nearly identical since the rail track has a stiff cross-section. Most importantly, the dominant frequency range of the vibrations aligns with the trackside noise, indicating a solid correlation between track vibration and rolling noise, and the potential to predict trackside noise based on track vibration.

3.3. Identification of pass-wheel time windows

The identification of the passing-wheel time windows can rely on the acceleration signal recorded on the rail tracks. In this case, the signal from the sensor A1 is employed. A recently coined method known as the three-threshold pulse extraction algorithm [57] is suitable for marking the ambit of the passing-wheel period. Fig. 8 presents the schematic diagram of the algorithm, and the detailed implementation procedure is presented as follows.

Step 1: Calculate the average energy of the acceleration signal (E_{mean}) as the reference threshold. The other two thresholds, the peak

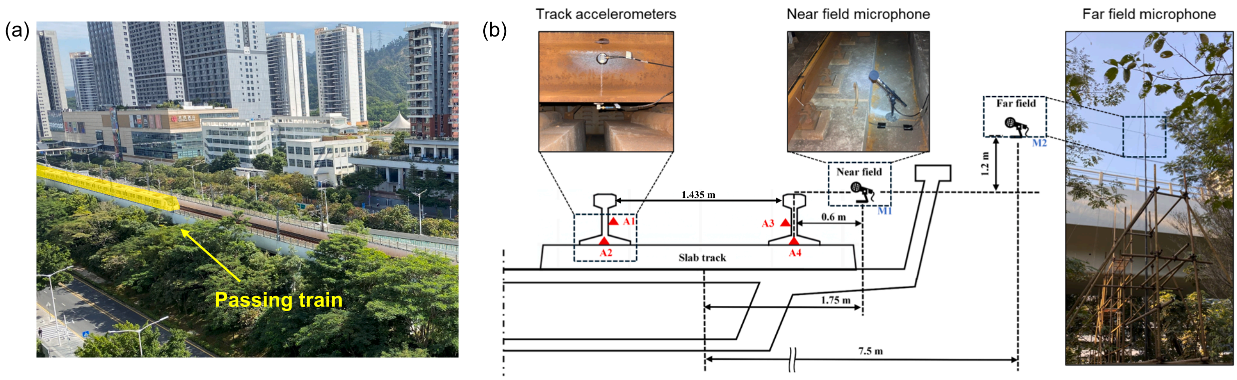

Fig. 5. (a) Environment at the experimental site; (b) Sensor arrangements on the test section.

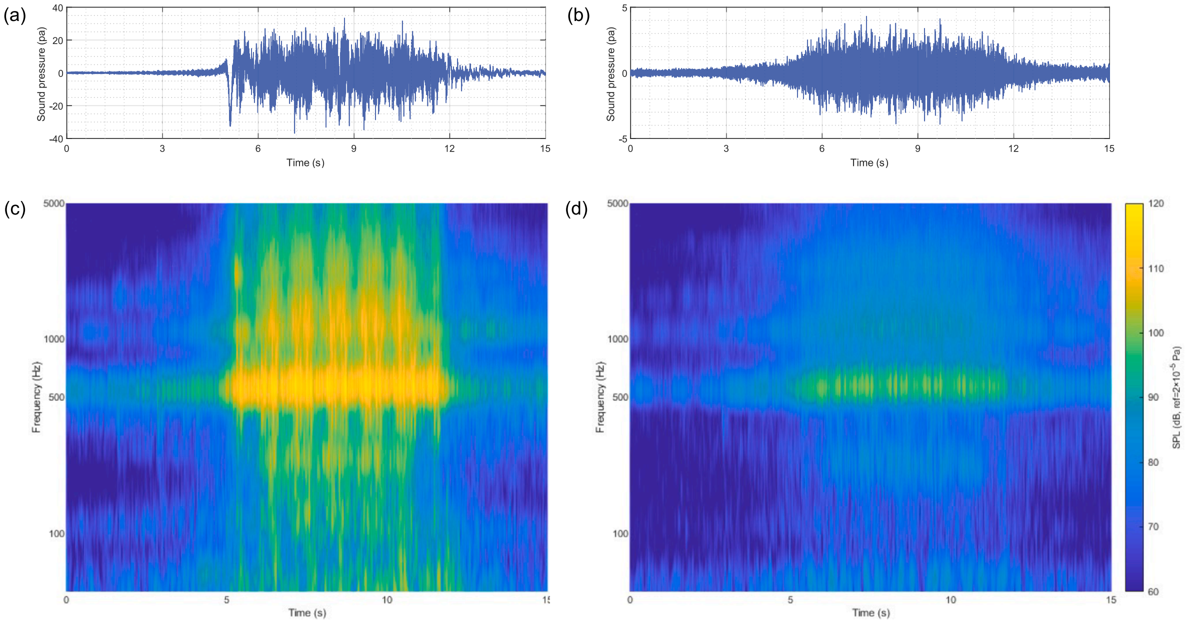

Fig. 6. Recorded trackside noise in time domain at the (a) near field, (b) far field; Recorded noise in time–frequency domain at the (c) near field, (d) far field.

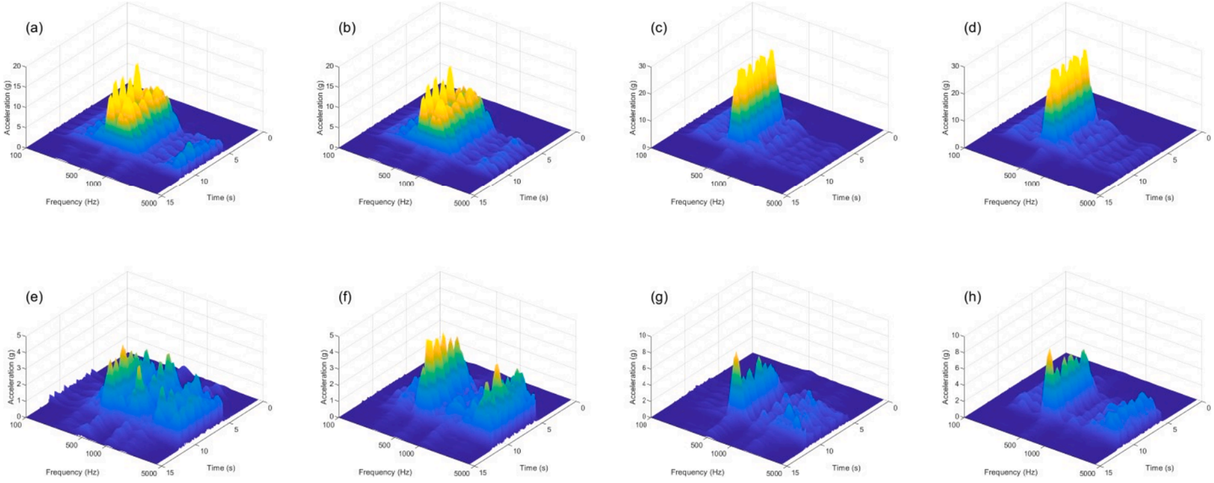

Fig. 7. Recorded track vibration of the channel (a) A1V, (b) A2V, (c) A3V, (d) A4V, (e) A1L, (f) A2L, (g) A3L and (h) A4L.

detection threshold (E_{peak}) and the endpoint threshold ($E_{criteria}$), are determined according to reference threshold. The coefficients used to calculate the thresholds are specially determined to follow the characteristics of different target signals. In this study, the peak threshold is calculated as $E_{peak} = 1.5E_{mean}$, endpoint threshold is set as $E_{criteria} = 1.2E_{mean}$.

Step 2: Conduct the first framing on the signal with the time window length of $\Delta t_1 = 200ms$, and calculate the energy of each frame. Search the time duration when the frame energy is higher than E_{peak}, and consider each duration as an effective pulse (nth passing-wheel period). For each effective pulse, find the frame with the largest energy, and mark the starting time of this frame as the peak time of the nth pulse t^n_{max}.

The operation metro train on the testing line has six cabins, each cabin is equipped with two bogies on both ends. Since the distance between the two end boogies of adjacent cabins is considerably small, these two boogies are regarded as a group to reflect an effective pulse (see Fig. 8). Therefore, n is equal to 7 in the first framing of the step 2.

Step 3: Conduct the second framing with $\Delta t_2 = 50ms$, and calculate the energy of each frame. Starting from the peak time of the nth pulse, search forward and backward until the frame energy is lower than $E_{criteria}$, and mark the end points as t^n_{start} and t^n_{end}.

Step 4: The duration of the nth pulse is given by t^n_{start} and t^n_{end}. Following this approach, identify all effective pulses by repeating the

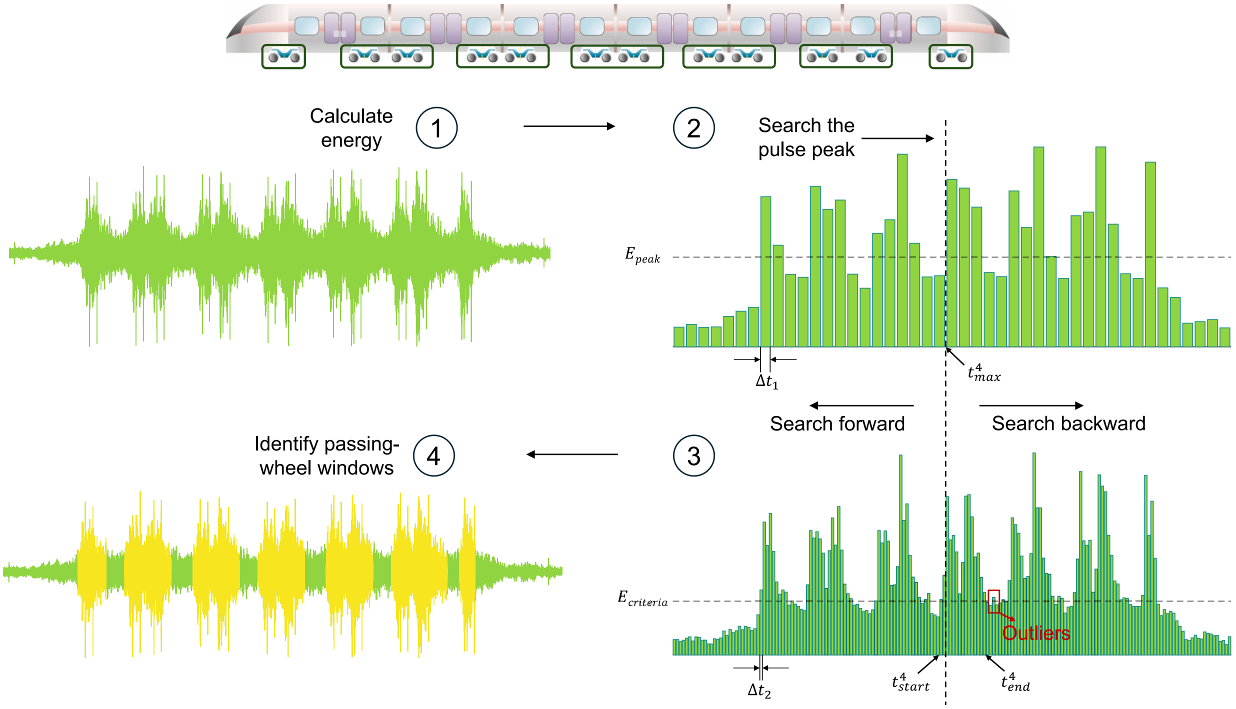

Fig. 8. Schematic diagram of the three-threshold pulse extraction algorithm.

previous steps, and review the identification results.

It is a more scientific approach for the identification of the passing-wheel time windows compared to setting a single threshold. For the latter approach, some outliers could be mistakenly identified as an effective pulse while actually no wheels are passing at that point.

4. Reconstruction results and Discussion

4.1. Reconstruction by the Echoformer

In this section, the reconstruction results by the Echoformer are compared with various baseline methods. The first baseline method is the ESN in the framework of RNN. Based on a standard ESN, a modified structure that replaces the linear readout layer with MLP is presented as the second baseline. This modified structure, named as ESDN, also possesses a recurrent structure for time series processing (Fig. 9). Another two methods that are suitable for sequential information decoding are introduced, including bi-directional LSTM [58] and vanilla Transformer [49].

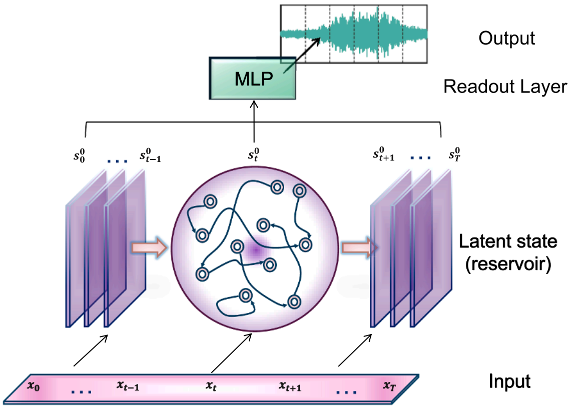

Fig. 9. Structure of the baseline method ESDN.

The hyperparameters of the proposed methods are confirmed by grid searching. The settings are listed below in Table 1. The size of the reservoir, N_s, is selected in $[100, 500, 1000]$. A larger value of N_s is restricted by the computational load. Other parameters related to the reservoir (i.e., $\boldsymbol{\rho}(\boldsymbol{W_{res}})$, $\boldsymbol{s}(\boldsymbol{W_{res}})$, and λ) are uniformly searched with the interval of 0.1, 0.5 and 0.1. The numbers of encoder and decoder blocks are set as 2, and a 10-head self-attention is adopted here. For the dataset separation, the 1st to the 4th passing-wheel windows (including periods without wheel passing, totally 3.65 s) are used for training, and the testing dataset starts from the end of the 4th window to the end of 7th window (length of 2.75 s). As shown in Fig. 6, the trackside noise level demonstrates a crescendo pattern at the beginning (4–6 s) and a decrescendo pattern toward the end (10–12 s). By separating these distinct patterns into the training and testing sets, the model's ability to uncover a more intrinsic relationship between vibration and noise can be evaluated. Another equally viable option is to train the model with the entire passing-train window, and validate the model with another passing train. The model training is conducted with the learning rate of 0.0001 for 50 epochs. An Intel Gold 5217 (CPU) and an NVIDIA Tesla P40 (GPU) are employed for the training.

An amendment should be made to the sound signal recorded at the far field. On account of the distance between the far field microphone to the rail track, i.e., the noise source, the sound propagates to the far field with a time delay. Take the midpoint between two rail tops as the hypothetical noise source (Fig. 10), the sound reaches the far filed microphone with a time delay of $\Delta t \approx 0.022$ s. Since the sampling rate of the microphone is 10000 Hz, the noise signal at the far field would be shifted for 200 points during training. The location of the near field microphone is close to the rail track, so no amendments are needed for the near field noise signal. It should be noted that this shift had little impact on the results due to the limited distance. However, for noise recorded over longer distances, incorporating the time delay factor is highly recommended.

Fig. 11 shows the overall reconstruction results on the entire testing dataset by the Echoformer. At the experimental site, the sound field near the rail track is more complex than the far field. Therefore, the amplitude of the sound pressure at the far field is relatively stable. It makes the reconstruction of the far field noise simpler, as observed in Fig. 11(b). As discussed in Section 3, the condition is divided into "with wheel" and "without wheel". Figs. 12 and 13 present a zoom in section in the middle of the signal at the near field with a length of 0.1 s under these two conditions.

It can be found that the sound energy with passing wheel is more concentrated and slightly increased. However, no significant differences can be concluded under the conditions presented in Figs. 12 and 13 as passing-wheel windows are too transient. For both conditions, the Echoformer successfully reconstructs the trend and spectrum of the signal. Based on the reconstructed noise signal, the sound pressure level (SPL) in the 1/3 octave band together with values of L_{eq} and L_{10}, which represents the overall and peak sound level [59], are calculated by all methods. With the sound pressure p, and the reference sound pressure $p_0 = 2 \times 10^{-5}$ pa, the calculation of SPL is given by:

$$SPL = 20\log_{10}(p/p_0). \tag{6}$$

SPL analysis results are presented in Figs. 14 and 15, and the reconstruction MSE for various methods are shown in Fig. 16. Suffering from the vanishing gradient problem, the bi-directional LSTM is unable to reach memories of long distances, it means that the LSTM is hindered in learning features of a long series data. Therefore, the LSTM fails in predicting the noise level both in the near and far field. The L_{eq} and L_{10} values given by LSTM are coarse and not acceptable. The other two RNN-based methods, ESN and ESDN, possess enhanced performance compared to the LSTM. The predicted noise results by the ESN and the ESDN have lower amplitude than the true values in the 1/3 octave band, and the prediction error increases with the frequency. Still and all, one can observe that these two methods manage to seize the overall trend of the target signal in the 1/3 octave band. Recall that the ESN is only equipped with a straightforward linear regression readout layer (Eq. (4)), the frequency features extraction ability should be owing to the high dimensional reservoir. Concluded from the results of all conditions, the performance of the ESN and the ESDN is indistinctive, meaning the replacement of the linear regression readout layer to the MLP is unnecessary. This fact is another proof for the effectiveness of the reservoir.

The Transformer structure can predict relatively accurate values of L_{eq} and L_{10}, meaning this method can be adopted for a rough analysis of the trackside noise amplitude. However, look into the results in the 1/3 octave band, one can find the recognition on the frequency features of the near field noise is a hurdle to the Transformer. Even for the far field noise recognition, the Transformer mistakenly raises the amplitude in the frequency range higher than 2000 Hz.

The proposed Echoformer merges the advantages of two methods, i.e., the frequency feature extraction ability of the reservoir, and the power of the Transformer to recognize the noise level. From the results, it is clear that the Echoformer delivers the trackside noise reconstruction with the highest quality, both in terms of frequency and amplitude. The specific prediction results on L_{eq} and L_{10} are listed in Tables 2 and 3.

Table 1
Hyperparameters used in the Echoformer.

Parameters	N_s	$\rho(\boldsymbol{W_{res}})$	$s(\boldsymbol{W_{res}})$	λ
Value	1000	0.8	2	0.1

Note: $\rho(\boldsymbol{W_{res}})$ is the spectral radius of $\boldsymbol{W_{res}}$, and the sparsity $s(\boldsymbol{W_{res}})$ is the ratio between L0 norm number and N_s.

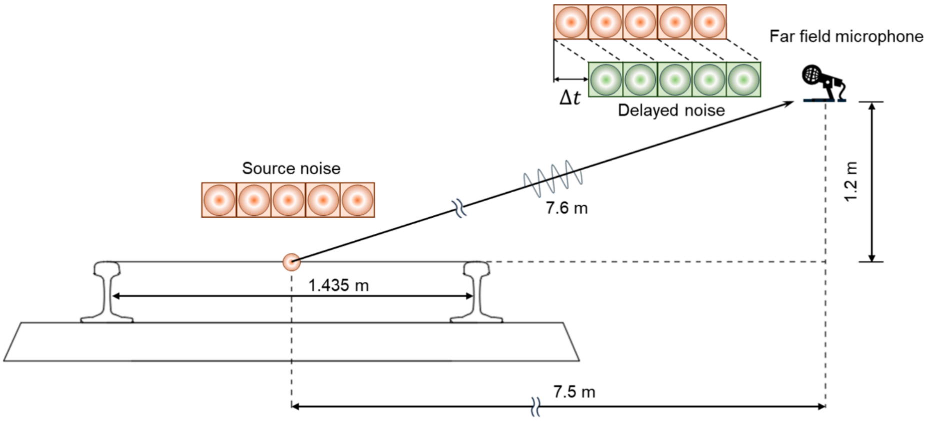

Fig. 10. Sound propagation to the far field microphone.

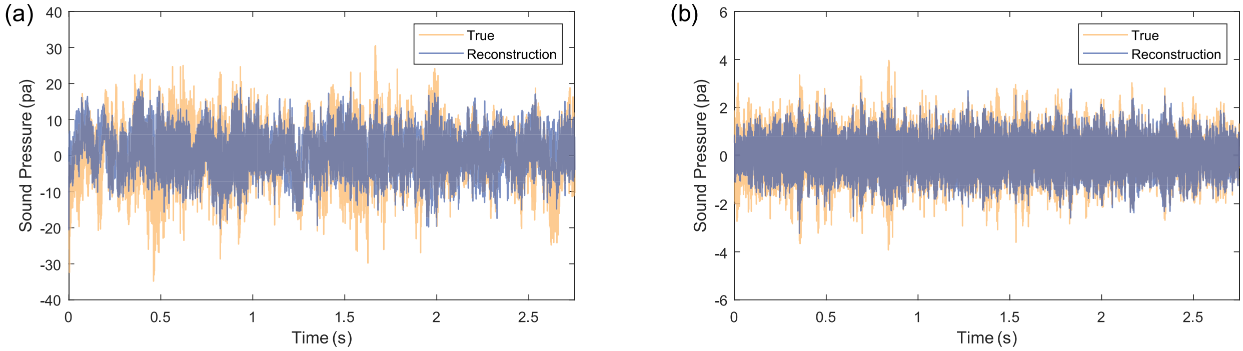

Fig. 11. Trackside noise reconstruction by the Echoformer at (a) near field and (b) far field.

4.2. Influence of missing input information

In the noise monitoring on railways, the integrity of signals corrected by sensors cannot always be promised. The provision of intact information to the Echoformer can theoretically guarantee the accuracy of model predictions. However, the robustness of the reconstruction should be examined when input information is partially missing. In this section, along with the original task, the signals from some of the accelerometers will be removed. All methods should give predictions with incomplete input information.

In the original task, four accelerometers (A1, A2, A3 and A4) records acceleration in two directions (V and L). Five input training set configurations will be compared in this section. The original task is marked as T1 with 8 input channels. Task T2 takes signals of accelerometers on the web of both rail tracks (A1 and A3). In task T3, sensors on the bottom of both rail tracks are counted. T4 and T5 get the smallest training set, which only obtains information from one sensor. No task that contains two sensors from a single rail track is investigated since the information on the rail web and the rail bottom is quite duplicated, see Fig. 7. Detailed information on different tasks is given in Table 4.

To facilitate the visualization on the performance of all models, the testing dataset is cut into slices of 0.05 s, that is 500 time steps for each slice counting the sampling rate of 10000 Hz. There will be a total of 55 slices on the testing dataset for both near and far fields, as the length of the testing dataset is 2.75 s. For the predicted trackside noise by all five methods, the mean square error (MSE) of all slices is calculated, and the distribution of the slices' MSE can manifest the performance of different methods, as well as the influence of various datasets. The mean value (μ) of the distribution is actually the MSE of one method on the entire testing dataset, and the standard deviation value (σ) shows the stability of the model reconstruction.

Fig. 17 presents the reconstruction results on the testing dataset. Apparently, the Echoformer has a superior performance across all methods in terms of accuracy. Yet, for the more challenging target of reconstructing near field noise, the standard deviation of the Echoformer is larger than the Transformer. The standard deviation of the ESN and the ESDN is also considerably large in the near field case. It means the MSE of some prediction slices generated by these methods is abnormally high (or low). These results may imply the imposition of the reservoir could sacrifice the stability of the reconstructed signal in time axis. However, it can also be seen that the MSE of the LSTM is lower compared to the ESN and the ESDN. Considering the fact that the performance of the LSTM is actually worse in this attempt, the MSE might not be the optimum index to evaluate the proposed methods. The reason is that MSE emphasizes deviations in the time domain, whereas the frequency domain is more critical for evaluating noise signals as shown in Figs. 14 and 15. Therefore, MSE can only be considered as a reference on the reconstruction performance.

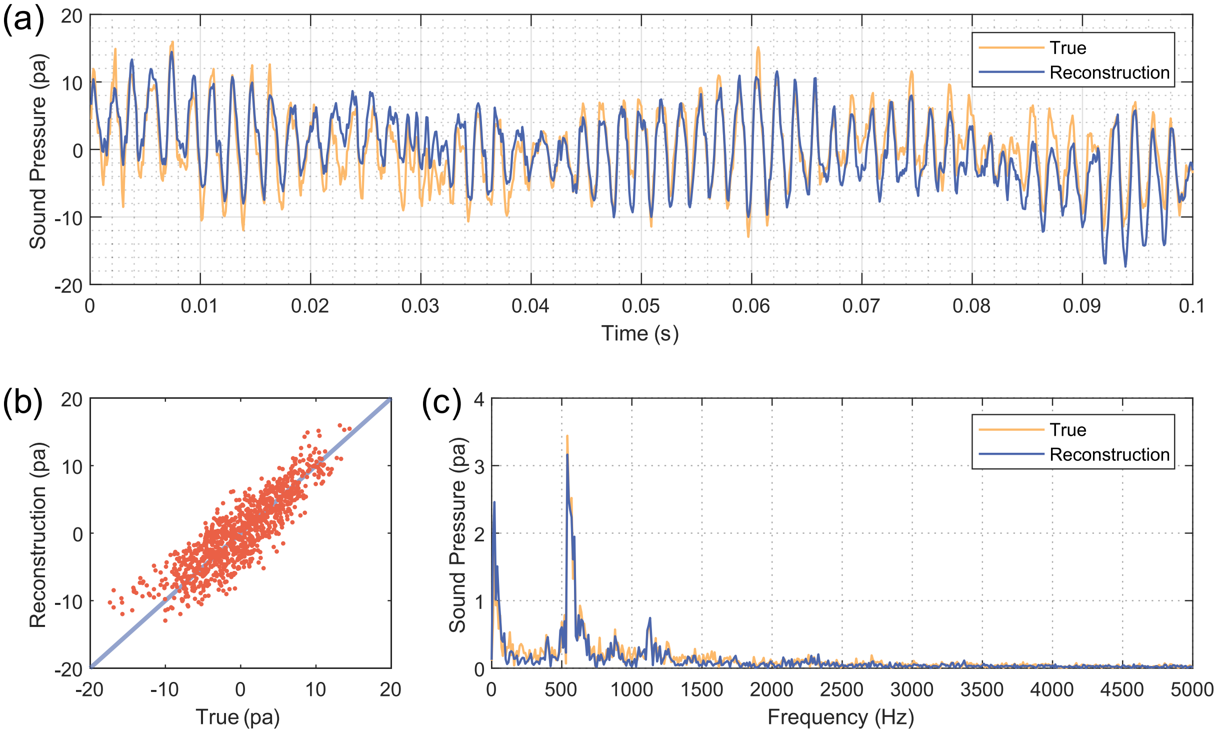

Fig. 12. A reconstructed section at the near field with wheels.

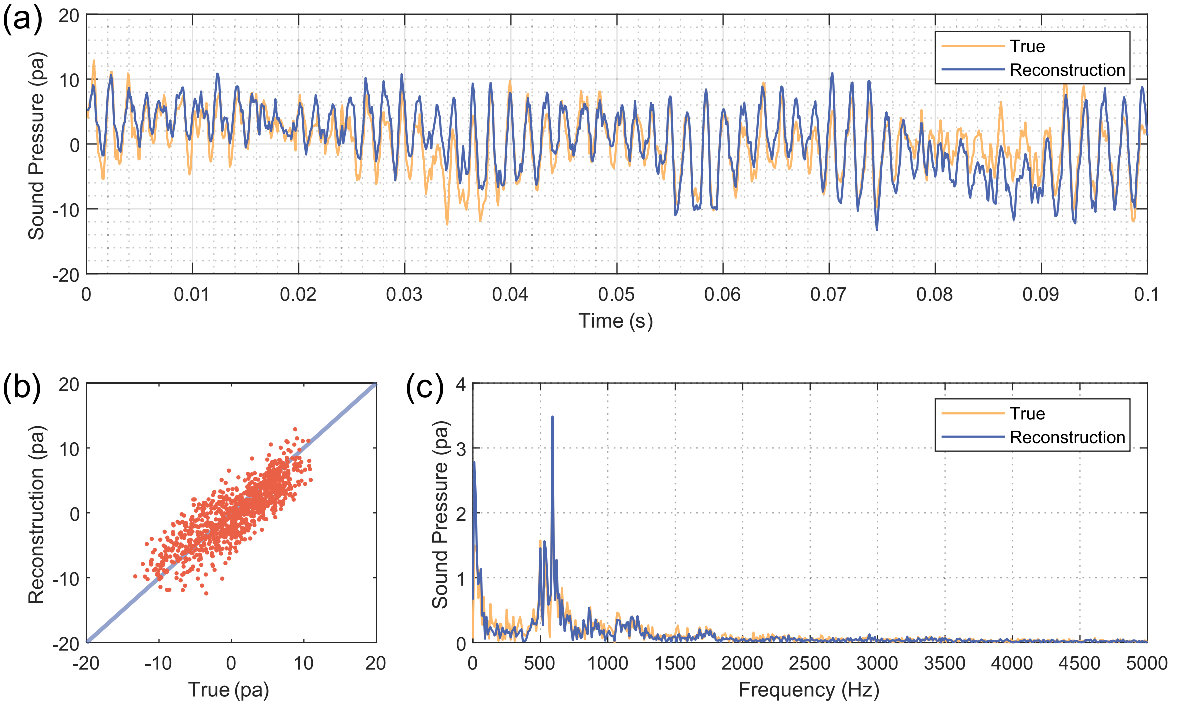

Fig. 13. A reconstructed section at the near field without wheels.

For the results on various tasks, no significant difference is presented by removing the information of sensors. The proposed method is still qualified for robust reconstruction with input signals from one accelerometer. The reason for this phenomenon could be the strong correlation between track vibration and railway rolling noise. Therefore, even one accelerometer can provide sufficient

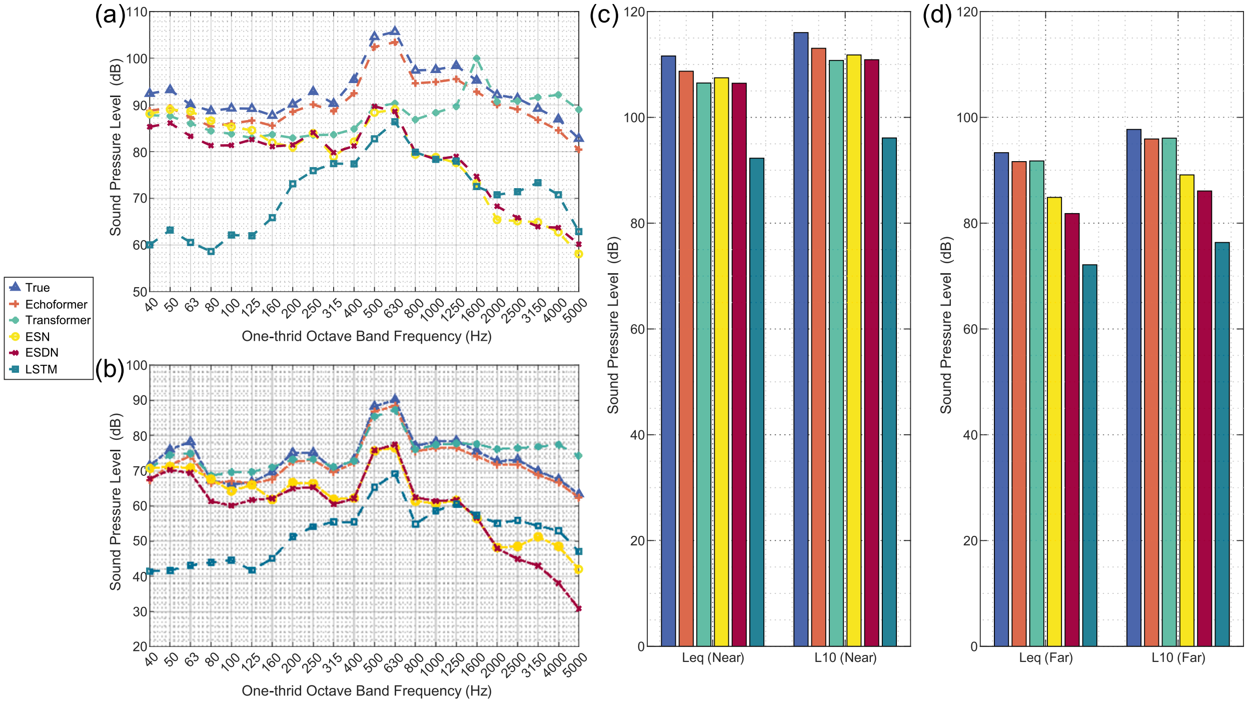

Fig. 14. SPL analysis with wheels in the 1/3 octave band at the (a) near field and (b) far field; L_{eq} and L_{10} values at the (c) near field and (d) far field.

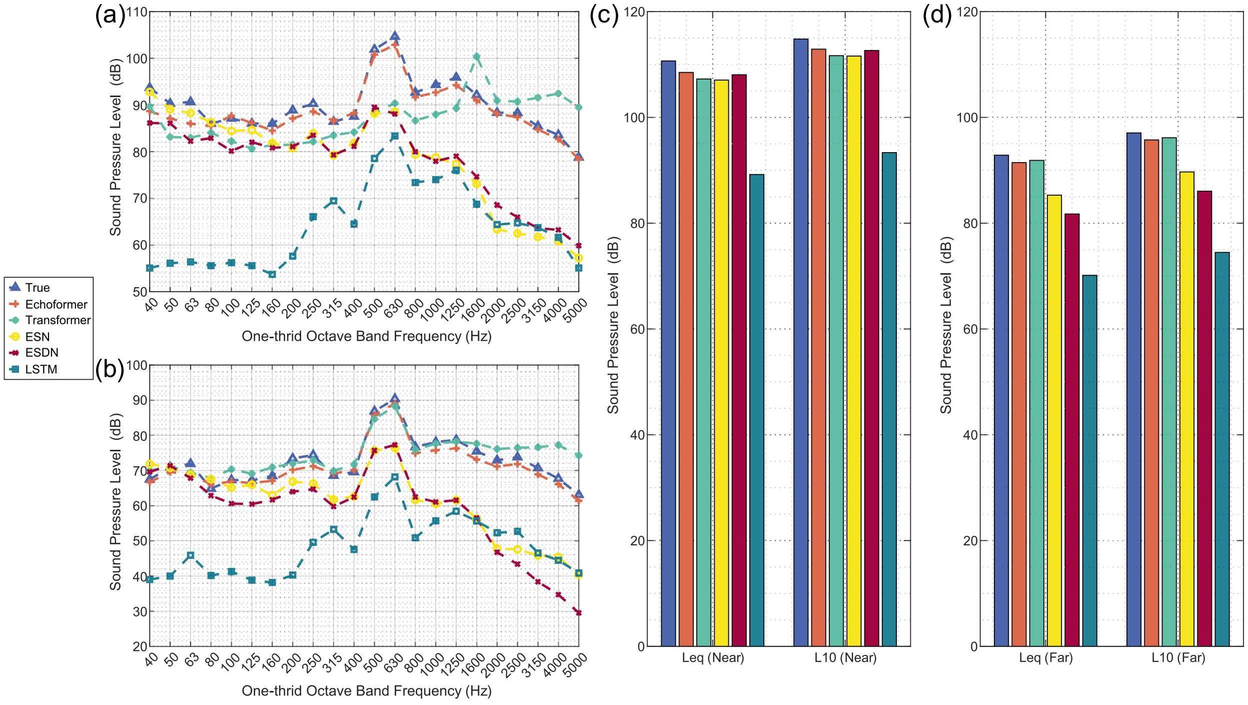

Fig. 15. SPL analysis without wheels in the 1/3 octave band at the (a) near field and (b) far field; L_{eq} and L_{10} values at the (c) near field and (d) far field.

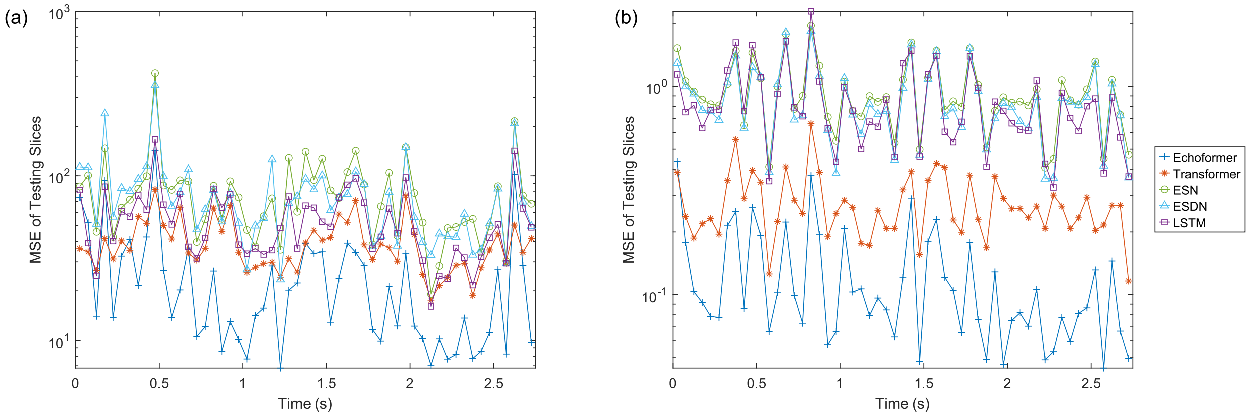

Fig. 16. Reconstruction MSE along time axis for various methods at the (a) near field and (b) far field.

Table 2
SPL analysis with wheels by all methods (dB).

SPL	True	Echoformer	Transformer	ESN	ESDN	LSTM
Near field						
L_{eq}	111.6	108.7	106.5	107.5	106.5	92.28
L_{10}	116.0	113.1	110.8	111.8	110.9	96.11
Far field						
L_{eq}	93.32	91.64	91.76	84.87	81.82	72.13
L_{10}	97.72	95.92	96.07	89.12	86.09	76.35

Table 3
SPL analysis without wheels by all methods (dB).

SPL	True	Echoformer	Transformer	ESN	ESDN	LSTM
Near field						
L_{eq}	110.7	108.5	107.3	107.1	108.1	89.21
L_{10}	114.8	112.9	111.7	111.6	112.6	93.34
Far field						
L_{eq}	92.85	91.45	91.87	85.30	81.73	70.15
L_{10}	97.06	95.75	96.15	89.69	86.05	74.48

Table 4
Input dataset information.

Input dataset	Known sensors	Missing sensors	Input channels
T1	A1, A2, A3, A4	/	8
T2	A1, A3	A2, A4	4
T3	A2, A4	A1, A3	4
T4	A2	A1, A3, A4	2
T5	A4	A1, A2, A3	2

information to rebuild the trackside noise. The average values of all testing slices' MSE by different methods are shown in Table 5.

4.3. Influence of noise-polluted input signal

Noise pollution in measurement data is inevitable in a monitoring project. In this section, the robustness of the Echoformer is investigated under the conditions when the signal input is disturbed by random white noise. Incremental levels of white noise will be added to the original training set of acceleration signals.

The white noise level is determined according to the energy of each input acceleration channel. The impact of white noise is studied with four levels, namely, 10 %, 30 %, 50 % and 70 % of the input of a single's energy. Fig. 18 (left) shows the original acceleration signal of channel A1V added with various levels of white noise. In this section, the model prediction by the Echoformer on the testing dataset is also cut into slices with the length of 500 time steps. Fig. 18 (right) presents the MSE of testing slices of the near field case

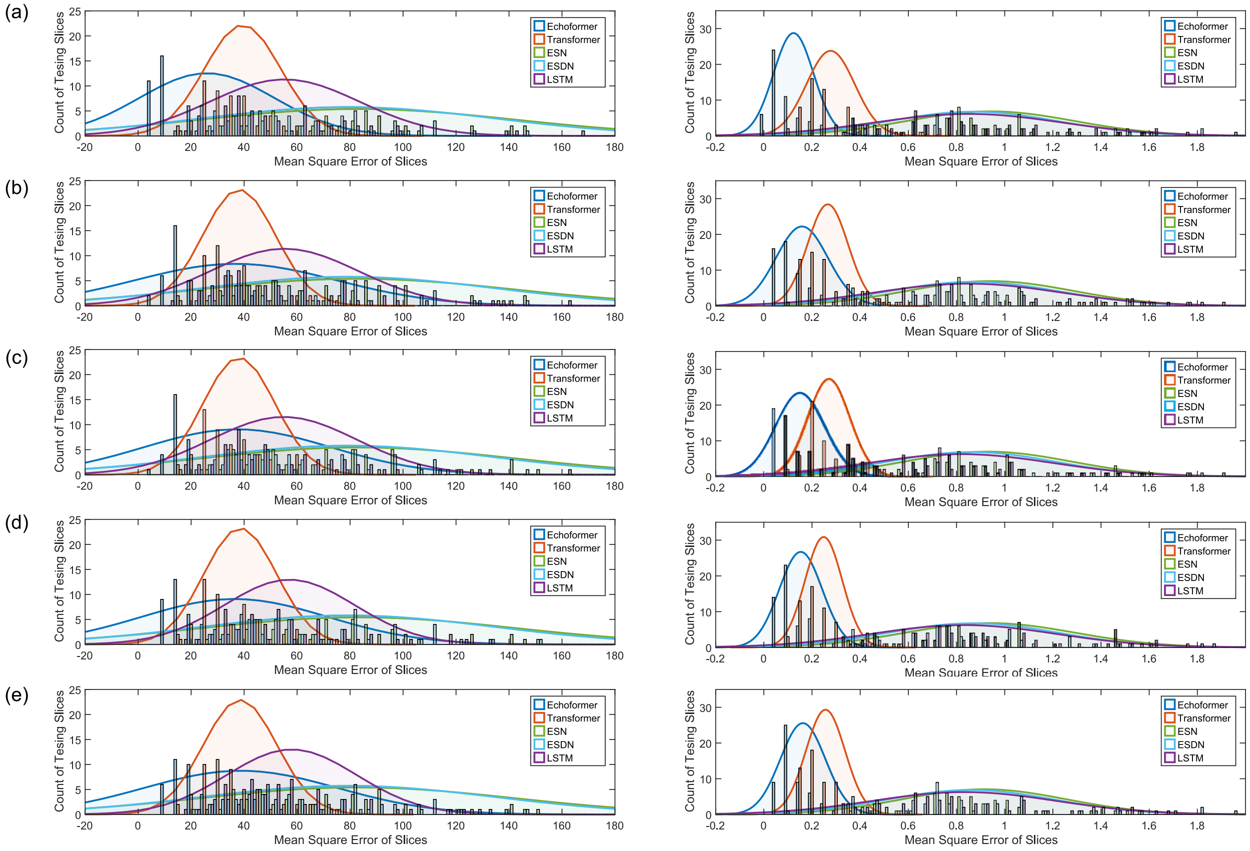

Fig. 17. Reconstruction results by different methods at the near field (left) and far field (right) on dataset (a) T1, (b) T2, (c) T3, (d) T4, and (e) T5.

Table 5
Average MSE of all testing slices by different methods.

Dataset	Echoformer	Transformer	ESN	ESDN	LSTM
Near field					
D1	25.95	39.43	82.77	79.31	55.30
D2	36.79	38.12	82.55	79.32	55.21
D3	37.44	38.09	82.97	79.19	54.93
D4	36.40	38.55	83.22	79.32	57.15
D5	38.74	38.51	82.71	79.23	57.96
Far field					
D1	0.123	0.277	0.941	0.869	0.850
D2	0.158	0.265	0.940	0.871	0.848
D3	0.150	0.270	0.931	0.861	0.818
D4	0.153	0.248	0.939	0.870	0.830
D5	0.162	0.256	0.915	0.858	0.833

along the time axis. It is evident that the influence of white noise is evenly distributed on the testing dataset, and the prediction error is roughly increasing with the noise level.

Again, the distribution of the testing slices' MSE is shown in Fig. 19. In general, the impact on the model prediction is not significant for both cases at near and far fields. The possible reasons are two-fold: first, the energy of the white noise is spread on the frequency range. Therefore, the disturbance on the specific frequency component is not high enough to influence the reservoir, which has already elucidated its ability in frequency feature extraction. Second, the self-attention block embedded in the Echoformer is intended to obtain the attention score, which is calculated through the SoftMax function. The essence of self-attention is the self-correlation within a piece of signal, and the original correlation in the input data is hard to change by noise. This algorithm is naturally insensitive to the noise, consequently, the Echoformer is endowed with the power to resist noise. In summary, the proposed method is robust in the situation when the input signal is polluted by noise.

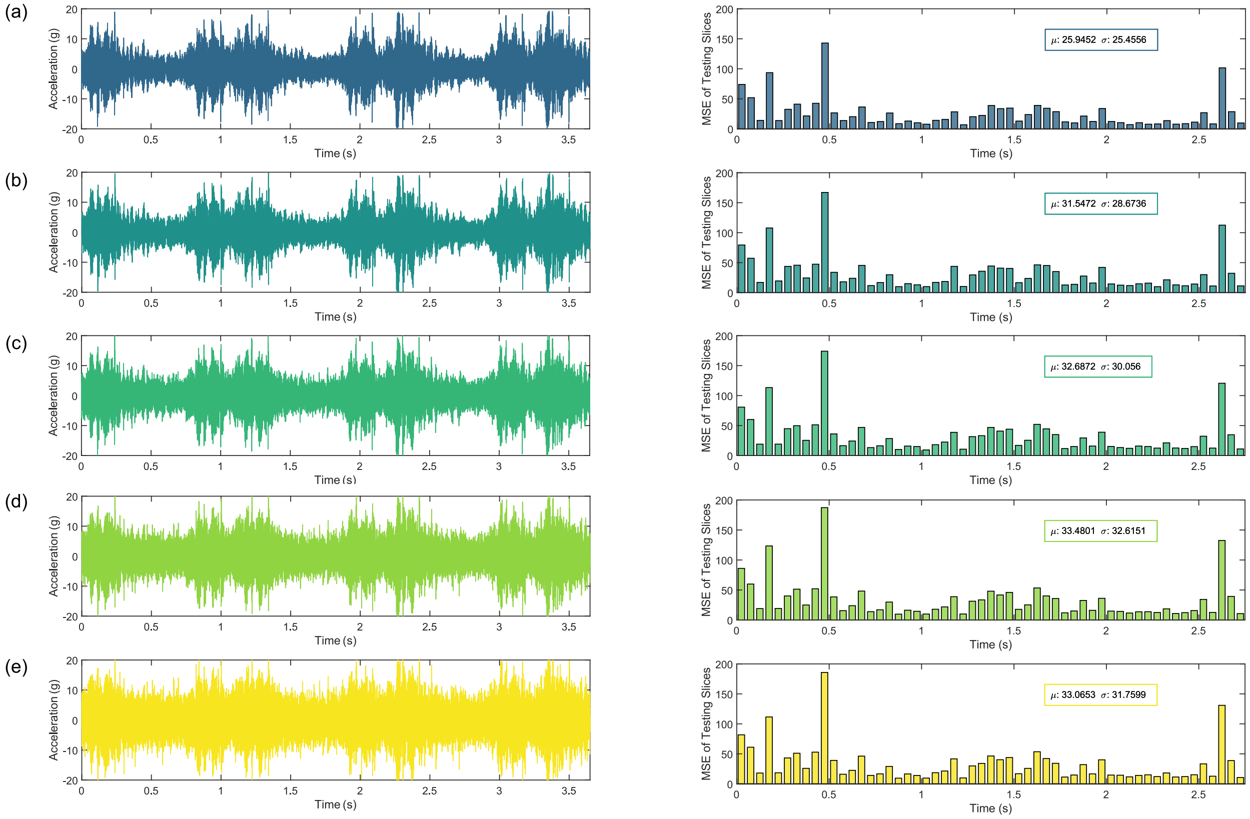

Fig. 18. Input signal adding incremental white noise (right), and the corresponding results (left) with noise level of (a) 0%, (b) 10%, (c) 30%, (d) 50%, and (e) 70%.

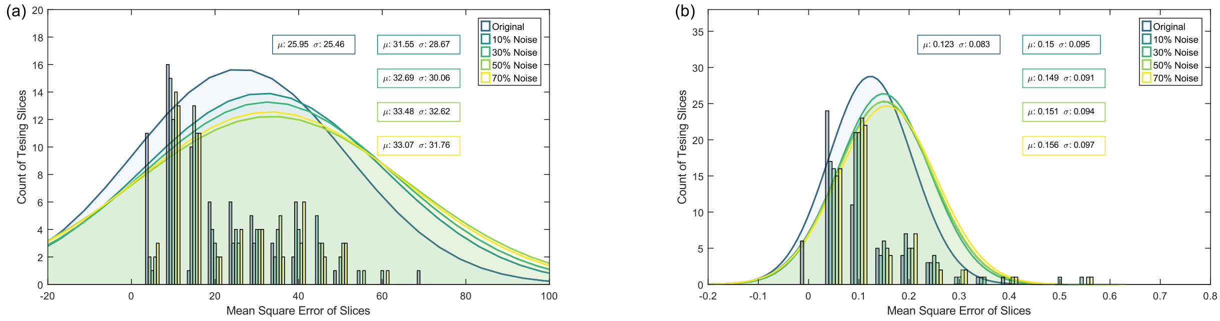

Fig. 19. Reconstruction results by the Echoformer under different noise levels at the (a) near field, and the (b) far field.

4.4. Model validation on another noise monitoring project

The Echoformer has proved its ability for a precise and robust reconstruction on the trackside noise from the track vibration data on a straight urban metro line. However, the previous work was done on a single metro railway, it will be plauded if the method has a certain generalization ability, which means it can be somehow adopted in other scenarios.

Aiming to test the generalization ability, the dataset stemmed from another monitoring project is used to verify the established surrogated mode. The project was conducted on the Wenzhou S1 metro line with similar conditions as the Shenzhen project: the measuring section was located at a straight viaduct line too, and the operating train (urban metro train type D, consistent with the general specifications of type A) also ran at the speed of 70 km/h.

In this project, only the far field microphone was set up for noise monitoring, which exactly fits the motivation of establishing the proposed surrogate model. However, the sensor deployment is different from the setup in the Shenzhen project. Four accelerometers were installed on the support and midspan of one rail track, and bottom sensor was attached to the rail flange, see Fig. 20(a). As a

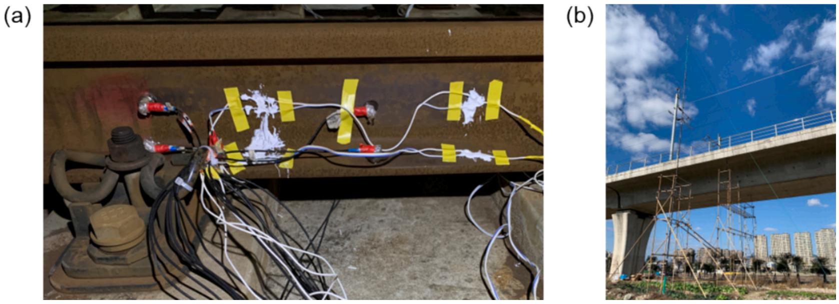

Fig. 20. (a) Sensor installation on the rail track, and the (b) far field microphone setup on the Wenzhou project.

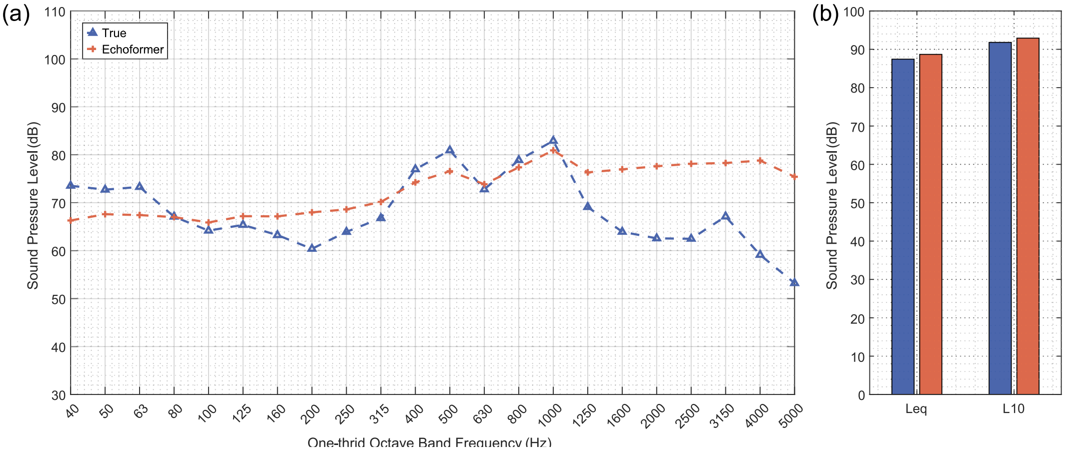

Fig. 21. Reconstruction results by the Echoformer in the (a) 1/3 octave band, and the predicted (b) L_{eq} and L_{10} values.

result, only one sensor can be used for the reconstruction of the trackside noise.

Fig. 21 shows the reconstruction results by the Echoformer in the Wenzhou project. It can be seen that the reconstruction is not ideal in another monitoring project, especially in the high-frequency range. However, the predicted noise has good agreement at the critical frequency range, i.e., from 400 Hz to 1000 Hz, with the true value. The prediction on the L_{eq} and L_{10} values can also roughly reflect the characteristics of the trackside noise. The true values of the L_{eq} and L_{10} are 87.43 dB and 91.80 dB, and the predicted values are 88.68 dB and 92.91 dB, respectively. Considering the complexity of the generation of the rolling noise, this prediction results are already satisfactory.

5. Conclusions and future works

In this work, a surrogate model that reconstructs trackside rolling noise based on the track vibration is established. This surrogate model adopts the RNN-Transformer structure, which has pronounced its state-of-the-art performance in relative works. Comparisons with various baseline methods reveal that the high-dimensional reservoir in the ESN is uniquely capable of recognizing the frequency characteristics of trackside noise. Besides, the Transformer structure is also proved to possess the superior ability in capturing the amplitude of the long series data, which is a challenge for RNN-based methods presented in this work. By combining the strengths of both approaches, the proposed Echoformer embeds the reservoir as the echo states to the Transformer to formulate the RNN-Transformer framework. In the reconstruction task, the echo state embedding first project the input signal, i.e., acceleration recorded on the rail track, to a sparse and high-dimensional latent space. This inputted features then go through the self-attention blocks to further establish the mapping between the trackside noise and track vibrations.

The Echoformer is amenable in reconstructing the rolling noise, delivering high-quality and robust predictions on the testing dataset, while also demonstrating strong generalization on the validation dataset. The key conclusions are summarized as follows:

(1) The proposed Echoformer outperforms the listed baseline methods, namely, the vanilla Transformer, the ESN and ESDN, and the bi-directional LSTM. The MSE of the predicted trackside noise by the Echoformer is 25.95 for the near field noise, and 0.123 for the far field noise.
(2) The Echoformer demonstrates robust reconstruction capabilities despite incomplete or noise-polluted data. When the input acceleration channels were reduced to a single sensor, the prediction MSE increased by 12.79 for the nearfield case and 0.039

for the far field case. For the noise-polluted scenario, with a 70 % noise level, the MSE increased by 7.12 for the near field and 0.033 for the far field.

(3) The generalization ability was validated using data from a separate noise monitoring project. The proposed Echoformer performed well in capturing noise levels within the critical frequency range of railway rolling noise. The errors in the predicted L_{eq} and L_{10} values are 1.25 dB and 1.11 dB.

In general, this work is an elementary attempt that proved the feasibility of reconstructing trackside noise through the track vibration. The developed surrogate model achieves high-quality predictions on the same urban metro line. However, its performance on other lines is less ideal. This is likely due to the fact that the generation of rolling noise is influenced by a wide range of factors. Even on the same line, rolling noise can vary over time as rail corrugation worsens due to wheel-rail contact. A more comprehensive surrogate model, capable of generalizing across different metro lines, could be developed by considering all relevant factors. At present, the established model is only suggested to be used on the original rail line which produces the training dataset, as it fulfills the two motivations of this work: compensate the monitoring data loss and facilitate the development of track vibration control devices. Furthermore, regarding the distinct mechanism of squeal noise on curve tracks, Echoformer requires further development to handle both straight and curved tracks simultaneously.

Regarding the modeling method, one notable drawback of Echoformer is its high memory consumption, as the reservoir significantly expands the dimensionality of the feature space. With the current training set, it requires nearly 30 GB of GPU memory. However, since the reservoir functions as a sparse feature extractor, a substantial portion of the feature space consists of zero values. Therefore, developing a more efficient representation of the reservoir is a highly desirable direction for future work.

CRediT authorship contribution statement

Xin Ye: Writing – original draft, Software, Methodology, Formal analysis, Conceptualization. **Yan-Ke Tan:** Writing – review & editing, Software, Methodology, Data curation, Conceptualization. **Yi-Qing Ni:** Writing – review & editing, Supervision, Project administration, 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.

Acknowledgements

The work described in this paper was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region (SAR), China (Grant No. PolyU 152308/22E). The authors would also like to appreciate the funding support by the Innovation and Technology Commission (ITC) of Hong Kong SAR Government to the Hong Kong Branch of Chinese National Rail Transit Electrification and Automation Engineering Technology Research Center (Grant No. K-BBY1 and ITS/096/21). The authors appreciate colleagues and students who participated in the noise monitoring projects, including Dr. Masoud Sajjadi, Dr. Yun-Ke Luo, Mr. Yu-Ling Wang, Mr. En-Ze Rui and Mr. Lei Yuan.

Data availability

Data will be made available on request.

References

[1] S. Weidenfeld, M.T. Schmitz, S. Sanok, A. Henning, D. Aeschbach, E.M. Elmenhorst, Effects of railway rolling noise on perceived pleasantness, Transp Res D Transp Environ 126 (2024), https://doi.org/10.1016/j.trd.2023.103995.

[2] M. Michali, A. Emrouznejad, A. Dehnokhalaji, B. Clegg, Noise-pollution efficiency analysis of European railways: A network DEA model, Transp Res D Transp Environ 98 (2021), https://doi.org/10.1016/j.trd.2021.102980.

[3] X. Deng, Z. Qian, Z. Li, R. Dollevoet, Investigation of the formation of corrugation-induced rail squats based on extensive field monitoring, Int J Fatigue 112 (2018) 94–105, https://doi.org/10.1016/J.IJFATIGUE.2018.03.002.

[4] D.J. Thompson, Railway noise and vibration mechanisms, modelling and means of control, 2013.

[5] J.C.O. Nielsen, A. Pieringer, D.J. Thompson, P.T. Torstensson, Wheel-Rail Impact Loads, Noise and Vibration: A Review of Excitation Mechanisms, Prediction Methods and Mitigation Measures, Notes on Numerical Fluid Mechanics and Multidisciplinary Design (2021), https://doi.org/10.1007/978-3-030-70289-2_1.

[6] BS EN 15461:2008+A1:2010. Railway applications-Noise emission-Characterization of the dynamic properties of track selections for pass by noise measurements; 2010., n.d. https://www.en-standard.eu/bs-en-15461-2008-a1-2010-railway-applications-noise-emission-characterization-of-the-dynamic-properties-of-track-selections-for-pass-by-noise-measurements/ (accessed March 27, 2022).

[7] J. Jin, H. Kim, H.I. Koh, J. Park, Railway noise reduction by periodic tuned particle impact damper with bounce and pitch-coupled vibration modes, Compos Struct 284 (2022), https://doi.org/10.1016/J.COMPSTRUCT.2022.115230.

[8] A. Pascale, C. Guarnaccia, E. Macedo, P. Fernandes, A.I. Miranda, S. Sargento, M.C. Coelho, Road traffic noise monitoring in a Smart City: Sensor and Model-Based approach, Transp Res D Transp Environ 125 (2023), https://doi.org/10.1016/j.trd.2023.103979.

[9] Y.K. Luo, L.Z. Song, C. Zhang, Y.Q. Ni, Experimental evaluation and numerical interpretation of various noise mitigation strategies for in-service elevated suburban rail, Measurement (lond) 219 (2023), https://doi.org/10.1016/j.measurement.2023.113276.

[10] Z. Zhang, Y. Luo, Restoring method for missing data of spatial structural stress monitoring based on correlation, Mech Syst Signal Process 91 (2017), https://doi.org/10.1016/j.ymssp.2017.01.018.

[11] H.P. Wan, Y.Q. Ni, Bayesian multi-task learning methodology for reconstruction of structural health monitoring data, Struct Health Monit 18 (2019), https://doi.org/10.1177/1475921718794953.

[12] K. Feng, J.C. Ji, Q. Ni, M. Beer, A review of vibration-based gear wear monitoring and prediction techniques, Mech Syst Signal Process 182 (2023), https://doi.org/10.1016/j.ymssp.2022.109605.

[13] Q. Ni, J.C. Ji, K. Feng, B. Halkon, A fault information-guided variational mode decomposition (FIVMD) method for rolling element bearings diagnosis, Mech Syst Signal Process 164 (2022), https://doi.org/10.1016/j.ymssp.2021.108216.

[14] Y. Yu, F. Han, Y. Bao, J. Ou, A Study on Data Loss Compensation of WiFi-Based Wireless Sensor Networks for Structural Health Monitoring, IEEE Sens J 16 (2016), https://doi.org/10.1109/JSEN.2015.2512846.

[15] W. Yang, S.K. Ahn, H. Koh, J. Park, Railway vibration reduction using impact dampers, INTERNOISE 2014 - 43rd International Congress on Noise Control Engineering: Improving the World Through Noise Control (2014) 3–5.

[16] G. Squicciarini, M.G.R. Toward, D.J. Thompson, Experimental procedures for testing the performance of rail dampers, J Sound Vib 359 (2015), https://doi.org/10.1016/j.jsv.2015.07.007.

[17] D. Gong, J. Zhou, W. Sun, Passive control of railway vehicle car body flexural vibration by means of underframe dampers, J. Mech. Sci. Technol. 31 (2017), https://doi.org/10.1007/s12206-017-0108-2.

[18] J. Jin, W. Yang, H.I. Koh, J. Park, Development of tuned particle impact damper for reduction of transient railway vibrations, Appl. Acoust. 169 (2020), https://doi.org/10.1016/J.APACOUST.2020.107487.

[19] W. Li, A. Wang, X. Gao, L. Ju, L. Liu, Development of multi-band tuned rail damper for rail vibration control, Appl. Acoust. 184 (2021), https://doi.org/10.1016/j.apacoust.2021.108370.

[20] L.Y. Liu, W.T. Cui, J.L. Qin, Q.M. Liu, L.Z. Song, Effects of rail pad viscoelasticity on vibration and structure-borne noise of railway box girder, Jiaotong Yunshu Gongcheng Xuebao/journal of Traffic and Transportation Engineering 21 (2021). https://doi.org/10.19818/j.cnki.1671-1637.2021.03.007.

[21] Z. He, Y. Bai, C. Su, N. Bao, P. Li, X. Zhang, L. Zhang, Y. Liu, Structural design of new mesh-type high-damping rail pad and comparative experimental study on the mechanical property with the traditional rail pads, Structures 57 (2023), https://doi.org/10.1016/j.istruc.2023.105234.

[22] Y. Zhang, X. Yang, S. Liu, Design and parameters influence analysis of dynamic vibration absorber for fastener clips in high-speed railway, Jvc/journal of Vibration and Control 30 (2024), https://doi.org/10.1177/10775463231154144.

[23] P. Wang, J. Lu, C. Zhao, L. Yao, X. Ming, Analysis on the Effects of Material Parameters on the Fatigue Performance of Novel Anticorrugation Elastic Rail Clips, Shock Vib. 2020 (2020), https://doi.org/10.1155/2020/5416267.

[24] Y.K. Luo, S.X. Chen, L. Zhou, Y.Q. Ni, Evaluating railway noise sources using distributed microphone array and graph neural networks, Transp Res D Transp Environ 107 (2022), https://doi.org/10.1016/j.trd.2022.103315.

[25] X. Ye, Y.Q. Ni, W.K. Ao, L. Yuan, Modeling of the hysteretic behavior of nonlinear particle damping by Fourier neural network with transfer learning, Mech Syst Signal Process 208 (2024), https://doi.org/10.1016/J.YMSSP.2023.111006.

[26] X. Ye, Y.-Q. Ni, M. Sajjadi, Y.-W. Wang, C.-S. Lin, Physics-guided, data-refined modeling of granular material-filled particle dampers by deep transfer learning, Mech Syst Signal Process 180 (2022), https://doi.org/10.1016/J.YMSSP.2022.109437.

[27] S. Wang, P. Perdikaris, Deep learning of free boundary and Stefan problems, J Comput Phys 428 (2020), https://doi.org/10.1016/j.jcp.2020.109914.

[28] Z. Ye, J. Yu, Deep morphological convolutional network for feature learning of vibration signals and its applications to gearbox fault diagnosis, Mech Syst Signal Process 161 (2021), https://doi.org/10.1016/J.YMSSP.2021.107984.

[29] M. Raissi, P. Perdikaris, G.E. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, J Comput Phys 378 (2019) 686–707, https://doi.org/10.1016/j.jcp.2018.10.045.

[30] K. Feng, J.C. Ji, Y. Zhang, Q. Ni, Z. Liu, M. Beer, Digital twin-driven intelligent assessment of gear surface degradation, Mech Syst Signal Process 186 (2023), https://doi.org/10.1016/j.ymssp.2022.109896.

[31] Q. Ni, J.C. Ji, B. Halkon, K. Feng, A.K. Nandi, Physics-Informed Residual Network (PIResNet) for rolling element bearing fault diagnostics, Mech Syst Signal Process 200 (2023), https://doi.org/10.1016/j.ymssp.2023.110544.

[32] S. Li, J. Ji, K. Feng, K. Zhang, Q. Ni, Y. Xu, Composite Neuro-Fuzzy System-Guided Cross-Modal Zero-Sample Diagnostic Framework Using Multi-Source Heterogeneous Non-Contact Sensing Data, IEEE Trans. Fuzzy Syst. (2024), https://doi.org/10.1109/TFUZZ.2024.3470960.

[33] H. Salehinejad, S. Sankar, J. Barfett, E. Colak, S. Valaee, Recent Advances in Recurrent Neural Networks, (2018).

[34] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning phrase representations using RNN encoder-decoder for statistical machine translation, in: In: EMNLP 2014–2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference, 2014, https://doi.org/10.3115/v1/d14-1179.

[35] G. Van Houdt, C. Mosquera, G. Nápoles, A review on the long short-term memory model, Artif Intell Rev 53 (2020), https://doi.org/10.1007/s10462-020-09838-1.

[36] J. Shan, X. Zhang, Y. Liu, C. Zhang, J. Zhou, Deformation prediction of large-scale civil structures using spatiotemporal clustering and empirical mode decomposition-based long short-term memory network, Autom Constr 158 (2024), https://doi.org/10.1016/j.autcon.2023.105222.

[37] R. Kabir, S.M. Remias, J. Waddell, D. Zhu, Time-Series fuel consumption prediction assessing delay impacts on energy using vehicular trajectory, Transp Res D Transp Environ 117 (2023), https://doi.org/10.1016/j.trd.2023.103678.

[38] H. Luo, M. Wang, P.K.Y. Wong, J. Tang, J.C.P. Cheng, Construction machine pose prediction considering historical motions and activity attributes using gated recurrent unit (GRU), Autom Constr 121 (2021), https://doi.org/10.1016/j.autcon.2020.103444.

[39] J. Man, H. Dong, X. Yang, Z. Meng, L. Jia, Y. Qin, G. Xin, GCG: Graph Convolutional network and gated recurrent unit method for high-speed train axle temperature forecasting, Mech Syst Signal Process 163 (2021), https://doi.org/10.1016/j.ymssp.2021.108102.

[40] P. Koprinkova-Hristova, I. Georgiev, M. Raykovska, Echo State Network for Features Extraction and Segmentation of Tomography Images, Comput. Sci. Inf. Syst. 21 (2024), https://doi.org/10.2298/CSIS230128045K.

[41] D. Yang, T. Li, Z. Guo, Q. Li, Multi-Scale Convolutional Echo State Network with an Effective Pre-Training Strategy for Solar Irradiance Forecasting, IEEE Access 12 (2024), https://doi.org/10.1109/ACCESS.2024.3349661.

[42] S. Bouazizi, H. Ltifi, Novel diversified echo state network for improved accuracy and explainability of EEG-based stroke prediction, Inf Syst 120 (2024), https://doi.org/10.1016/j.is.2023.102317.

[43] H. Jaeger, H. Haas, Harnessing Nonlinearity: Predicting Chaotic Systems and Saving Energy in Wireless Communication, Science 304 (2004) (1979) 78–80, https://doi.org/10.1126/SCIENCE.1091277/SUPPL_FILE/JAEGER_SOM.PDF.

[44] M. Lukoševičius, H. Jaeger, B. Schrauwen, Reservoir Computing Trends, KI - Kunstliche Intelligenz 26 (2012) 365–371, https://doi.org/10.1007/S13218-012-0204-5/FIGURES/1.

[45] K. Yeo, Data-driven reconstruction of nonlinear dynamics from sparse observation, J Comput Phys 395 (2019) 671–689, https://doi.org/10.1016/J.JCP.2019.06.039.

[46] E. Haugsdal, E. Aune, M. Ruocco, Persistence Initialization: a novel adaptation of the Transformer architecture for time series forecasting, Appl. Intell. 53 (2023), https://doi.org/10.1007/s10489-023-04927-4.

[47] B. Wu, C. Fang, Z. Yao, Y. Tu, Y. Chen, Decompose Auto-Transformer Time Series Anomaly Detection for Network Management †, Electronics (switzerland) 12 (2023) https://doi.org/10.3390/electronics12020354.

[48] Q. Wen, T. Zhou, C. Zhang, W. Chen, Z. Ma, J. Yan, L. Sun, Transformers in Time Series: A Survey, in: IJCAI International Joint Conference on Artificial Intelligence, 2023. https://doi.org/10.24963/ijcai.2023/759.

[49] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, in: Adv Neural Inf Process Syst, 2017.

[50] P. Fan, G. Song, Z. Zhai, Y. Wu, L. Yu, Fuel consumption estimation in heavy-duty trucks: Integrating vehicle weight into deep-learning frameworks, Transp Res D Transp Environ 130 (2024), https://doi.org/10.1016/j.trd.2024.104157.

[51] S. Li, J.C. Ji, Y. Xu, K. Feng, K. Zhang, J. Feng, M. Beer, Q. Ni, Y. Wang, Dconformer: A denoising convolutional transformer with joint learning strategy for intelligent diagnosis of bearing faults, Mech Syst Signal Process 210 (2024), https://doi.org/10.1016/j.ymssp.2024.111142.

[52] M. Bani-Almarjeh, M.B. Kurdy, Arabic abstractive text summarization using RNN-based and transformer-based architectures, Inf Process Manag 60 (2023), https://doi.org/10.1016/j.ipm.2022.103227.

[53] S. Liu, S. Zhang, X. Zhang, H. Wang, R-Trans: RNN Transformer Network for Chinese Machine Reading Comprehension, IEEE Access 7 (2019), https://doi.org/10.1109/ACCESS.2019.2901547.

[54] R. Xia, M. Zhang, Z. Ding, RTHN: A RNN-transformer hierarchical network for emotion cause extraction, in: IJCAI International Joint Conference on Artificial Intelligence, 2019. https://doi.org/10.24963/ijcai.2019/734.

[55] M. Lukoševičius, H. Jaeger, Reservoir computing approaches to recurrent neural network training, Comput Sci Rev 3 (2009) 127–149, https://doi.org/10.1016/J.COSREV.2009.03.005.

[56] R. Xiong, Y. Yang, D. He, K. Zheng, S. Zheng, C. Xing, H. Zhang, Y. Lan, L. Wang, T.Y. Liu, On layer normalization in the transformer architecture, in: 37th International Conference on Machine Learning, ICML 2020, 2020.

[57] W. Tao, Z. Sun, G. Wang, S. Xiao, B. Liang, M. Zhang, S. Song, Broiler sound signal filtering method based on improved wavelet denoising and effective pulse extraction, Comput Electron Agric 221 (2024), https://doi.org/10.1016/j.compag.2024.108948.

[58] B. Jang, M. Kim, G. Harerimana, S.U. Kang, J.W. Kim, Bi-LSTM model to increase accuracy in text classification: Combining word2vec CNN and attention mechanism, Applied Sciences (switzerland) 10 (2020), https://doi.org/10.3390/app10175841.

[59] S. Namba, S. Kuwano, T. Kato, An investigation of L eq, L 10, and L 50 in relation to loudness, J Acoust Soc Am 64 (1978), https://doi.org/10.1121/1.2004282.