Journal of Civil Structural Health Monitoring (2025) 15:2887–2901
https://doi.org/10.1007/s13349-025-00938-6

ORIGINAL PAPER

Bayesian optimization bidirectional LSTM approach for the condition assessment of underground-operating trains

Yu-Ling Wang[1,2] · Yuhang Lu[1,2] · Yan-Ke Tan[1,2,3] · Wai Kei Ao[1,2] · Yi-Qing Ni[1,2] · Qing-Chen Tang[1,2]

Received: 1 May 2024 / Accepted: 7 March 2025 / Published online: 2 May 2025

Abstract

Underground railways play a crucial role in global intercity transport networks, necessitating the implementation of diverse measures to mitigate vibrations during train operations. However, with the variable damping, the structures of underground trains can inadvertently impact passenger's comfort when taking them. Consequently, the development of the online monitoring system becomes imperative to assess the operational conditions of these trains. This research applies the ISO2631 standard to analyze the dynamic responses of train's accelerations, utilizes the ride comfort index to determine the operational state of the train, and uses online monitoring data to evaluate its overall conditions. The study proposes an online monitoring system that utilizes the long short-term memory (LSTM) algorithm, which has demonstrated effectiveness in time-series prediction and identification tasks. By learning from historical and future signal segments, the LSTM algorithm enables the diagnosis and identification of underground train-operating conditions under varying working conditions. To enhance the accuracy of prediction results, the algorithm is optimized by adopting the bi-directional structure and Bayesian optimization method. Quantitative analyses demonstrate that the optimized bi-directional LSTM model achieves a correlation up to 94.32% for overall dataset and 90.45% on test dataset. Finally, an illustrative case is presented to highlight the performance of the proposed method in assessing the conditions of underground trains.

Keywords Rail vibration control · Rail transit structural health monitoring · Online condition assessment · Train vibration time-series analysis · Bidirectional LSTM · Bayesian optimization

✉ Wai Kei Ao
waikei.ao@polyu.edu.hk

Yu-Ling Wang
owen-yl.wang@polyu.edu.hk

Yuhang Lu
yuhang.lu@polyu.edu.hk

Yan-Ke Tan
yanke.tan@connect.polyu.hk

Yi-Qing Ni
ceyqni@polyu.edu.hk

Qing-Chen Tang
cheney.tang@polyu.edu.hk

[1] Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China

[2] National Rail Transit Electrification and Automation Engineering Technology Research Center (Hong Kong Branch), Hung Hom, Kowloon, Hong Kong, China

[3] College of Civil Engineering, Tongji University, No. 1239 Siping Rd., Yangpu District, Shanghai, China

1 Introduction

Underground railways have become increasingly prevalent in major cities around the globe due to their ability to accommodate large amounts of passengers, their efficiency, and their positive environmental impact. These railways are frequently planned and implemented in densely populated areas, encompassing commercial, public, and residential spaces, with the aim at minimizing the commuting time and maximize the productivity.

However, alongside the convenience brought by rail transit systems, there have also emerged challenges related to vibration and noise, which have a direct impact on the daily lives of individuals in these areas. It was reported that vibrations coming from wheel-track interactions typically were transferred through the infrastructures to neighboring buildings, which could lead to abnormal building vibrations [1]. Concurrently, vibrations induced by rail irregularities, and vehicle dynamic and aerodynamic instability reveal an ineluctable issue for underground trains [2, 3]. The consequent

abnormal vibration can significantly influence the train's operation and passenger's comfort. Numerous vibration reduction techniques are applied to the infrastructure. However, some of these may influence the train's operation and the body vibration of a train running, even when structural vibration is reduced. Ultimately, these factors can impact passenger's comfort. Hence, the online condition assessment and monitoring of trains are pivotal in guaranteeing the safe and efficient maintenance of urban transportation during their operations.

The acceleration time-series of a train body typically are used as a crucial parameter for evaluating train operating conditions [4]. Besides, numerous structural health monitoring (SHM) methods are employed to assess the corresponding passenger's comfort [5]. Broadly, these methods are divided into two categories: model-driven and data-driven approaches. In the model-driven paradigm, train's operation is assessed by developing a finite element(FE) model of both the train and track, incorporating principles of structural mechanics, and examining the wheel-rail interaction [6]. However, the FE method has its own disadvantages, such as modeling complexity, lengthy computation times, and also the established FE models are of limited applicability. Moreover, uncertainties may arise from the experimental data used in the course of establishing a model, due to variations in experimental conditions and environmental discrepancies during the testing, potentially lowering the accuracy of the model.

In recent years, there has been a notable increase in the utilization of data-driven techniques for health monitoring of the structural system, which rely exclusively on the system response and environmental monitoring data [7]. A Bayesian modeling approach that incorporates model class selection for dynamic forecasting was developed, and applied to the real-world monitoring data collected from a three-tower cable-stayed bridge [8]. An overview of a structural performance monitoring system implemented on a butterfly-shaped arch footbridge located at The Hong Kong Polytechnic University was provided [9]. Two machine learning algorithms were compared to detect structural changes using statistical features derived from the raw dynamic data [10]. This approach was evaluated through experimental tests conducted on a laboratory beam structure and an actual railway bridge in France. An unsupervised, automated, data-driven strategy for SHM focused on detecting damage in railway bridges based on the system's dynamic response was proposed [11], which enhanced high sensitivity to subtle structural changes by extracting damage-sensitive features from the train-induced response when detecting anomalies. A long-term modal behavior of a footbridge was racked, utilizing low-cost sensors for frequency extraction and employing data-driven models to predict structural health [12]. A deep learning-based fault diagnosis framework was proposed, which is robust against label noise advancing the application of data-driven techniques in SHM [13]. Several studies have employed data-driven methods to enhance information extraction from low signal-to-noise acceleration data. For instance, the optimization of SSI-COV input parameters improved SHM accuracy by maximizing data extraction [14]. A low-cost SHM system for short-span bridges was also developed using data-driven approaches [15, 16]. Moreover, a novel modal testing method using synchronized data loggers enhanced real-time SHM efficiency and accuracy [17–20]. Lastly, data-driven techniques were emphasized to address the uncertainties in frequency extraction from time-domain signals, improving overall frequency analysis [21]. However, it required offline data processing, which limited the real-time monitoring and early warning capabilities. Additionally, the method was sensitive to environmental and operational factors, potentially leading to false or overlooked damage detection.

Within the realm of data-driven methods for SHM, recurrent neural networks (RNNs) are commonly used to address time-series correlation challenges [22]. RNNs, initially introduced by Hopfield et al. [23], are a preferred architecture for the sequential data. Unlike conventional artificial neural networks (ANNs), which consist of an input layer, a hidden layer, and an output layer connected by weights, RNNs incorporate conventional feed forward ANNs where outputs are fed back as inputs. This characteristic enables RNNs to excel in capturing temporal dependencies, rendering them well-suited for the time-series data analysis. However, RNNs often face difficulties in handling problems with long-term dependencies due to the potential instability of network training caused by abrupt increases in training weights, commonly referred to as exploding gradients. To address the challenges associated with long-term dependencies, long short-term memory(LSTM), a variant of RNN, was introduced [24]. LSTM is specifically designed to mitigate the vanishing gradient issue through its gating mechanisms, which regulate the flow of information and allow the network to maintain long-term dependencies. While LSTM is highly effective at handling the vanishing gradient problem, it may still encounter exploding gradients under certain conditions, which can be managed with techniques like gradient clipping [25]. LSTM offers significant advantages in processing long time-series data compared to traditional RNN methods [26], exhibiting superior accuracy and fast processing efficiency making it a powerful tool for managing and analyzing time-series data with long-term dependencies.

Recently, there has been a notable surge in integration of LSTM algorithms with structural health monitoring (SHM) methods for processing the experimental data [27]. The LSTM framework was employed to detect damage states

of a bridge using its deflection data gathered from SHM by tracking characteristic changes in the time-series deflection and temperature data [28]. Additionally, LSTM methods have been incorporated into SHM systems for short-term wind speed prediction, which helps improve bridge traffic control during severe wind conditions, especially by enhancing the accuracy of control time limits during typhoons [29]. In another approach, SHM data from bridges were processed using both Kohonen neural networks and LSTM networks, with the LSTM demonstrating superior predictive capabilities, providing timely alerts for structural changes [30]. Furthermore, BiLSTM networks have been combined with Generative Adversarial Networks (GANs) to recover missing or corrupted SHM data in large-scale bridges, proving effective in restoring time-series data [31]. LSTM also plays a vital role in predicting the evolving structural conditions over time, offering valuable insights into the long-term health of bridges [32]. Moreover, some researchers have integrated LSTM autoencoders with multi-head self-attention mechanisms to model time-series data, thereby improving feature extraction in unsupervised vibration-based damage diagnosis [33]. A notable example of LSTM's application is its combination with a multi-Gaussian fitting feature extraction method, which was developed as part of an SHM system utilizing vibration signals. Experimental results showed that this approach significantly reduced computation time while maintaining high prediction accuracy [34]. Despite LSTM's effectiveness in time-series prediction, it faces the common challenge of hyperparameter optimization, a difficulty shared with other neural networks. Key hyperparameters such as the number of LSTM layers, hidden layers, initial learning rate, and regression rate need to be carefully tuned to optimize the algorithm's performance. Techniques like Bayesian optimization have been employed to address this issue, improving the efficiency and accuracy of LSTM predictions in SHM applications.

Accordingly, this research aims to derive an online evaluation technique for train operating conditions and passenger's comfort, achieved by analyzing the acceleration time-series of trains operating under varying slab track conditions. First, the Bayesian parameter optimization method is employed for searching LSTM hyperparameters, which significantly enhances the efficiency by considering the algorithm's prediction accuracy, thus improving the overall efficiency of the algorithm, and the prediction error as a measure is also utilized to quantify the variation in train operating condition characteristics, as observed through the BiLSTM algorithm. Additionally, the comfort index factor a_w from the ISO2631-1:1997 [35] standard is introduced as a quantifiable metric for assessing passenger's comfort, the statistical data of different rail-track slab plates during their operations from trains running on a fixed schedule, route, speed, and batch type to minimize the interference from other factors. Finally, the SHM data from an underground train line serves as a case study to evaluate the method's effectiveness, and the effect of the Bayesian para-meter optimization method on the accuracy of LSTM prediction time-series is also examined.

2 Methodology

2.1 Architecture of the LSTM model

Figure 1 illustrates the structure of model proposed in this study. The mapping association does not require a complex architecture; thus, the algorithm consists of an input layer, hidden layers, decoder layers, and an output layer. Besides, a methodology proposed by researchers [36] is introduced to the design of this structure, which has been refined through a rigorous process of experimentation.

Figure 1 also denotes five cells of the hidden layer where the internal structure of each cell is explained in greater detail. Each cell comprises three gates: the forget gate f_t - responsible for determining which information should be removed from the memory; the input gate I_t, determining which information should be added to the memory; and the output gate O_t, determining which information from the memory should be retrieved. The subindex t refers to the time step. Additionally, the cell is characterized by two states: the cell state C_t and the hidden state h_t, all of which retain their values over arbitrary time intervals, while the three gates regulate the influx and efflux of information of these states.

Specifically, both the cell state C_t and hidden state ht have a dimension of 60 and are initialized as $C_0 = 0$ and $h_0 = 0$, based on empirical findings, representational needs, and computational efficiency considerations. The dimension of the hidden state, a critical hyperparameter

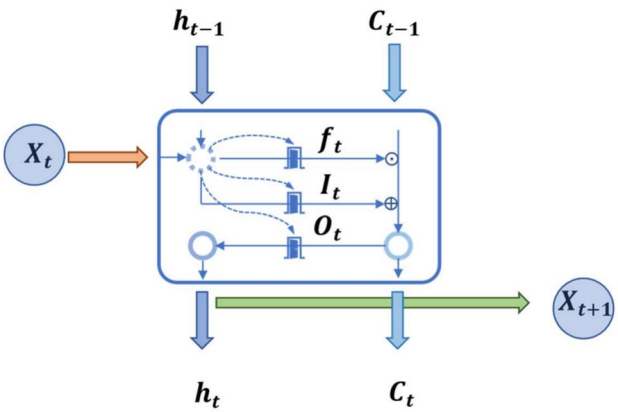


Fig. 1 Architecture of one LSTM hidden layer cell

in machine learning models, is typically adjusted based on the model's performance against a validation dataset. This parameter choice is indicative of the complexity of the data representations that the model is capable of learning. Specifically, a larger dimension facilitates the learning of more intricate representations, but it may also precipitate overfitting if the model's complexity is disproportionate to the size of the training data. Furthermore, computational efficiency can be a factor, as certain dimensions may enable more efficient computations depending on the specific hardware and calculation methods employed. Previous research has experimented with various dimensions and found that a dimension of 60 yielded the best performance [37]. Therefore, selecting a dimension of 60 is likely to strike a balance between capturing the complexity of the data and mitigating the risk of overfitting.

The primary role of the forget gate is to selectively eliminate information, acting as a mechanism to discard irrelevant data from the previous cell states. This is achieved by multiplying the prior cell states with a filter. Hence, the mathematical expression representing the forget gate, denoted as f_t, is as follows:

$$f_t = \sigma\left(W_f x_t + W_f h_{t-1} + b_f\right) \tag{1}$$

where σ is a sigmoid function to calculate the activation values ft of the forget gates at time step t; The weight parameter W_f is associated with the forget gate; The input data for this cell x_t consists of the comfort index factor a_w, which is a key parameter for assessing passenger comfort based on vibration data, derived from the ISO 2631-1:1997 standard, quantifies the level of vibration that passengers experience and is crucial for evaluating the operational conditions of the train system [35]. h_{t-1} is the hidden state transmitted from the preceding cell in the training stage. In contrast to the forget gate, the purpose of the input gate i_t is to enhance the information within C_t, which is mathematically expressed as follows:

$$i_t = \sigma\left(W_i x_t + W_i h_{t-1} + b_i\right) \tag{2}$$

where W_i symbolizes the weight parameters of the input gate; b_i is the bias parameters of the input gate.

Then, the LSTM layer selects the data that should be incorporated into the cell states C_t of the network. This process consists of the following three steps:

1. Calculate the candidate values $\hat{C}_t$ at the time step t, representing potential additions to the cell states, which can be mathematically represented as follows:

$$\hat{C}_t = \tanh\left(W_c x_t + W_c h_{t-1}\right) \tag{3}$$

where W_c is the weight parameters of the candidate vector.

2. Determine the activation values C_t for the input gates, multiplying it to $\hat{C}_t$, storing the information in the cell state, which can be expressed using the following equation:

$$C_t = f_t C_{t-1} + i_t \hat{C}_t \tag{4}$$

3. Obtain the output of memory cells within the LSTM layer o_t through the utilization of the following equations:

$$o_t = \sigma\left(W_o x_t + W_o h_{t-1} + b_0\right) \tag{5}$$

where W_0 symbolizes the weight parameters of the output gate; b_0 is the bias parameters of the output gate. The hidden state h_t, serves a dual purpose: transmitted to the subsequent cell and representing the output of the current cell, which can be performed as follows:

$$h_t = o_t \tanh\left(C_t\right) \tag{6}$$

After passing through all five LSTM cells, each cell generates an output h_t.

In this algorithm, the focus of predictions is the value of the next time step, which means that only h_5 is considered as the output from one hidden layer. As illustrated in Fig. 2, a decoder layer is employed to convert the multidimensional output h_n, with the dimension of n, which means each output has one sequence, with the length of the sequence n, and thus the total dimension of the input and hidden cells is 60. The output of the LSTM model y is calculated as follows:

$$y = Wh_n + b \tag{7}$$

where W corresponds to the weight parameters of the decoder layer, necessitating optimization in the training stage. Similarly, the bias parameters of the decoder layer, represented by b, also require optimization throughout the training process.

2.2 Bidirectional long short term memory (BiLSTM)

BiLSTM networks are characterized as an advanced extension of conventional LSTMs, highlighting their ability to incorporate information from both the past and future states, leading to an improved prediction accuracy [38]. Unlike conventional LSTMs that solely consider past observations and process information in an unidirectional manner, BiLSTM networks process the data in both forward and backward directions by training two separate LSTM networks, where the first one as a forward LSTM processes the in-put sequence forward, and the second one

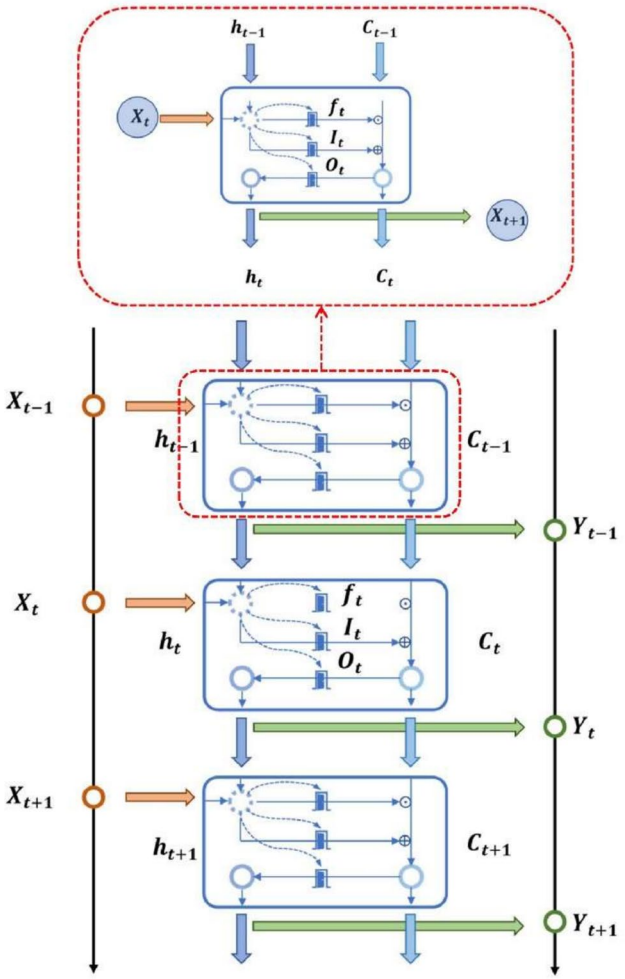


Fig. 2 Flow chart of standard LSTM network

as a backward LSTM processes a reversed duplication of the input sequence (the former is from beginning to end, but the latter is in reverse). The final prediction is then based on the outputs of both LSTMs combining information from both directions.

Thus, this concept was introduced to process the data in both directions by employing two distinct hidden layers [39], which were subsequently forwarded to the same output layer. As illustrated in Fig. 3, a BiLSTM computes the forward hidden sequence $\mathbf{h}_t$, and the backward hidden sequence $\overleftarrow{h}_t$.

$$\mathbf{h}_t = H\left(W_{1\overrightarrow{h}} x_t + W_{2\overrightarrow{h}} \mathbf{h}_{t-1} + b_{\overrightarrow{h}}\right) \tag{8}$$

$$\overleftarrow{h}_t = H\left(W_{3\overleftarrow{h}} x_t + W_{5\overleftarrow{h}} \mathbf{h}_{t+1} + b_{\overleftarrow{h}}\right) \tag{9}$$

where $H(\bullet)$ is the forward or backward LSTM layers of the BiLSTM networks, respectively.

Finally, the output sequence y is obtained by iteratively processing the backward layer from $t = T$ to $t = 1$ and the

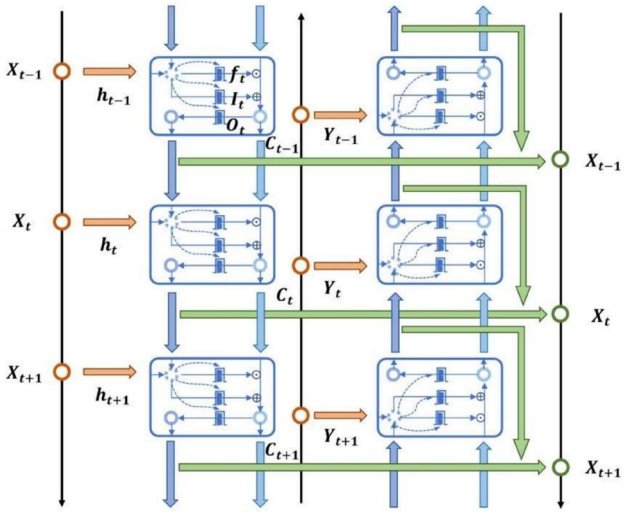


Fig. 3 A BiLSTM network

forward layer from $t = 1$ to $t = T$. The output layer is then updated as follows:

$$y_t = W_{4\overrightarrow{h}} \mathbf{h}_t + W_{6\overleftarrow{h}} \overleftarrow{h}_t + b_y \tag{10}$$

2.3 Bayesian-based BiLSTM parameters optimization

In the context of optimization, a prevalent strategy involves converting a problem into an optimization task and utilizing a variety of techniques. However, for machine learning algorithms, their capacity to align with both the model and the data are impacted by a set of hyperparameters. These hyperparameters are established prior to the learning process and play an essential role within the optimization function.

In a manner similar to other NNs, it is feasible to stack and connect BiLSTM layers. The quantity of BiLSTM layers and the number of units within each layer significantly impact both the training process accuracy and the model's capacity. This capacity, in turn, governs the model's ability to adapt to its intended purpose. To optimize model performance, it is advisable to avoid excessively high or low capacity. A model with insufficient capacity may struggle to adequately learn from the training data, while an overly complex model risks overfitting to the specific characteristics of the training dataset. Consequently, determining of these two hyperparameters becomes crucial in achieving an optimal model performance.

Similarly to the majority of deep learning algorithms, the BiLSTM algorithm involves multiple hyperparameters that impact various aspects of its functionality, including computational efficiency and the ability to generate accurate predictions. However, manually tuning these

hyperparameters is inefficient and prone to inaccuracies, as it requires a substantial amount of time for researchers to comprehend the influence of each hyperparameter on the algorithm and the overall generalization of the model. Hence, it becomes crucial to automate the search for optimal hyperparameters. Previous research has shown that Bayesian optimization is well-suited for identifying optimal hyperparameters in the context of BiLSTM, particularly when dealing with complex problems that involve long-duration differentials or expressions [40].

In this research, the specific attention is given to the following five hyperparameters: number of layers; number of units; number of BiLSTM layer(s); initial learning rate and regularization. The selection of these five hyperparameters for optimization is based on their critical influence on the performance, complexity, and generalization ability of the BiLSTM model when handling time-series data. First, the number of layers and number of units directly affect the model's complexity and expressive power. A few layers may lead to underfitting, while too many layers could result in overfitting. The number of BiLSTM layers is particularly important because BiLSTM networks can capture both forward and backward dependencies in time-series data, which is crucial in SHM applications. The initial learning rate is another key factor that affects the speed and efficiency of model training. Optimizing the learning rate helps the model converge more quickly without overshooting the optimal solution. Finally, regularization controls the overfitting risk by penalizing large weights in the model, ensuring that it generalizes well to unseen data. Exploring a range of regularization values helps maintain a good balance between fitting the training data and ensuring robust performance on new data. Compared to other possible hyperparameters, these five have a more direct and significant impact on the BiLSTM model's ability to process time-series data effectively, making them the primary focus for optimization. The optimization range of values for these hyperparameters is presented in Table 1, which is crucial for the value selection and should combine with theoretical insights, empirical validations, and computational feasibility.

The optimization range for these hyperparameters is designed to maintain an appropriate level of model complexity and prevent overfitting:

Table 1 Hyperparameters and the range of values

Hyperparameters	Ranges of values
Number of layers	[1, 4]
Number of units	[50, 300]
Number of BiLSTM layer(s)	[1, 2]
Initial learning rate	$[10^{-3}, 1]$
Regularization	$[10^{-10}, 10^{-2}]$

1. The number of layers is restricted to between [1, 4] to ensure sufficient model complexity.
2. The number of neural units is set between [50, 300] to capture important patterns in data while avoiding an over-parameterized model that could overfit.
3. The number of BiLSTM layers is kept between [1, 2] to benefit from bidirectional sequence processing.
4. Initial learning rates are chosen from [0.001, 1] to balance efficient convergence and the risk of overshooting the optimal solutions.
5. Regularization parameters range from $[10^{-10}, 10^{-2}]$ to provide a spectrum of weight penalties while preserving model complexity.

During the Bayesian optimization process, it can be assumed that the mean square error (MSE) loss function of the parameter-fitting model follows a Gaussian process. By modifying parameters, the posterior probability of this function is obtained. Thus, the optimization process can be divided into the following five steps:

1. The collective hyperparameters $[h1, h2,..., h5]$ are selected to calculate the loss function $[f(h1), f(h2),..., f(h5)]$, which follows the Gaussian distribution, the loss function of which can be expressed as follows:

$$f\left(h_{1:5}\right) \sim N(0, K)$$
$$K = \begin{bmatrix} k\left(h_1, h_1\right) \cdots k\left(h_1, h_1\right) \\ \ddots \\ k\left(h_5, h_1\right) \cdots k\left(h_5, h_5\right) \end{bmatrix} \tag{11}$$

   where K is defined as the core function and represents the covariance of the above distribution. Unlike other optimization methods like gradient descent algorithm, Bayesian optimization does not require an explicit expression of the function f. This step initializes the process by sampling from the parameter space to estimate the model's performance.
2. The expected improvement (EI) function is employed in the selection of sample points by estimating the potential improvement achievable when exploring the vicinity of the current optimal value. If the observed improvement after executing an algorithm falls below the expected level, it suggests that the current optimal point may be susceptible to a local optimum [41]. In such situations, the algorithm aims to identify an optimal point in another region within the given domain. The present $h+1$ is selected as the optimal combined parameters, with y^*, greater than $(h+1)$, as the threshold value. Furthermore, the three-structured Parzen estimator (TPE) is chosen to formulate the EI function [42], which can be mathematically expressed as follows:

$$\begin{cases} EI_{y^*}(h) = \int_{-\infty}^{y^*} (y^* - y) h(y|h) \mathrm{d}y = \frac{\gamma y^* \ell(h) - \ell(h) \int_{-\infty}^{y^*} h(y) \mathrm{d}y}{\gamma \ell(h) + (1-\gamma) g(h)} \\ h(h|y) = \begin{cases} \ell(h) \text{if} y \le y^* \\ g(h) \text{if} y \ge y^* \end{cases} \end{cases} \tag{12}$$

where, $\ell(h)$ is the distribution function which is smaller than y^*, $g(h)$ is the distribution function which is greater than y^*.

3. Then, the selection of a new sampling point involves a comparison of the loss function of the existing point. A next set of hyperparameters is selected by optimizing EI function, which balances exploration (trying new hyperparameter combinations) and exploitation (focusing on regions that are likely to yield good results). The new sampling point is identified as the one that maximizes the value of the EI function.

$$h_{\text{next}} = \arg\max_{p} EI(h) \tag{13}$$

4. To update the Gaussian process model $f(x)$ and derive a posterior distribution over functions, the new Gaussian distribution is obtained by incorporating the next sampling point as follows:

$$\begin{cases} \begin{bmatrix} f(h_{1:n}) \\ f(h_{next}) \end{bmatrix} \sim N\left(0, \begin{bmatrix} K & k' \\ k'T & k(h_{next}, h_{next}) \end{bmatrix}\right) \\ k' = \left[k(h_{\text{next}}, h_1), \cdots, k(h_{\text{next}}, h_n)\right] \end{cases} \tag{14}$$

The model is trained with the newly selected hyperparameters, and the corresponding loss value is computed. This new information is then used to update the Gaussian process model, improving its predictions for future iterations. And the Gaussian distribution of $f(h_{n+1})$ is shown as follows:

$$\begin{cases} f(h_{n+1})|f(h_{1:n}) \sim N(\mu, \sigma^2) \\ \mu = k'^T K^{-1} f(h_{1:n}) \\ \sigma^2 = k(h_{\text{next}}, h_{\text{next}}) - k'^T K^{-1} k' \end{cases} \tag{15}$$

5. Steps 2 and 3 are repeated in an iterative manner until a predetermined stopping criterion is met. This criterion may involve reaching the maximum number of iteration steps or ensuring that the difference between the current and best values does not exceed a specified threshold.

The algorithm can be clearly divided into two components: the updating of the posterior distribution and the optimization of the acquisition function. As additional observations are integrated, the posterior distribution is continually updated. Utilizing the updated posterior distribution, the algorithm identifies points that maximize the acquisition function and includes them in the training data set.

2.4 Detection algorithm: Bayesian factor and three-sigma rule

Within the domain of Bayesian statistics, Bayesian factor is introduced as an intuitive and principled approach for model selection [43], and a quantitative measure that compares the relative likelihood (the strength of evidence) of the observed data under two competing models. This allows for the support of one model over the other based on the obtained evidence.

When conducting a comparison between two models, model 1 (M1) and model 2 (M2), characterized by their parameter θ, are assessed in relation to the data D in eq. 16:

$$b_{10} = \frac{p(D|M_1)}{p(D|M_2)} = \frac{\int p(D|\theta_1, M_1) p(\theta_1|M_1) d\theta_1}{\int p(D|\theta_2, M_2) p(\theta_2|M_2) d\theta_2} \tag{16}$$

From the equation provided, it is evident that the Bayesian factor b_{10} represents the ratio between two marginal likelihoods. In this context, it explains the correlation between the two models by considering the overall goodness of each model to the observed data and the influence of the prior probabilities assigned to the model parameters. Essentially, the Bayesian factor serves as a scaling factor that quantifies the degree to which the probability of a model is modified, either increasing or decreasing, as a result of new evidence from the data, denoted as D. This adjustment takes into account the change in probability regardless of the prior probability assigned to supporting model 1.

One prevalent interpretation (as shown in Table 2) of the Bayesian factor was initially proposed in 1961 [44] and later enhanced in 2022 [45].

Moreover, another alternative form is shown as follows:

$$\frac{P(M_1|D)}{P(M_2|D)} = \frac{P(M_1)}{P(M_2)} b_{10} \tag{17}$$

Equation 17 can be interpreted as "Posterior model odds = Prior model odds $\times b_{10}$", focusing on the impact of the observed data on the likelihood of a model for the case study in Sect. 3.

Table 2 Interpretation scheme for values of the Bayesian factor

	Interpretation	Bayesian factor
M1	Very strong support	10 to 30
	Strong support	3 to 30
	Weak support	1 to 3
	No support for either model	1
M2	Weak support	0.33 to 1
	Positive support	0.1 to 0.33
	Strong support	0.033 to 0.1

The identification of the turning point in this time series involves the utilization of the three-sigma limit theorem as a suitable threshold for the minimum acceptable loss [46]. A control range of 0.27 % is employed to define the parameter range for subsequent calculations.

In practice, sigma serves as a statistical measure of variability, indicating the degree to which the observed data deviates from the mean value. During statistical calculations, the data points are evaluated to determine if they fall within three standard deviations (sigma) from the mean value. Typically, data points are expected to follow a distribution centered around the mean value and within predefined limits. In a well-controlled process, approximately 99.73 % of the data is anticipated to fall within plus or minus three standard deviations (sigma) from the mean value. For instance, in a normal distribution, less than 1 % of the data points lie beyond the three-sigma threshold, either above or below the mean value. Therefore, if the difference between the predicted value and the actual value surpasses this threshold, it can be considered a significant turning point or outlier.

3 Case study

3.1 Instrumentation system

For the purpose of this site test, a section of the traditional slab (TS) track between two neighboring stations has been replaced with a floating slab track (FST) infrastructure in the underground railway system. This replacement was aimed at mitigating structural vibrations within the underground transit conduits. To assess the effectiveness of this infrastructure replacement in reducing vibrations, a comparison was made between the vibrations exhibited by trains passing over the newly installed FST track and the conventional TS track. A specific underground line was selected to minimize external interference and ensure data consistency. The line had fixed operational parameters-consistent train schedules, routes, and speeds-allowing us to control variables and focus on the effects of the track change. The only modification was the replacement of the traditional slab track (TS) with the floating slab track (FST), while all other conditions, such as train type, speed, and time of day, remained constant. This setup isolated the track change's impact on vibration and passenger comfort, ensuring that any performance differences were solely due to the track type change.

To conduct this evaluation, a sensing device was installed within the driver's cab of the train. This device consisted of strategically placed piezoelectric accelerometers positioned on the cab floor to measure accelerations. During the entire operational period, this system continuously gathered real-time monitoring data.

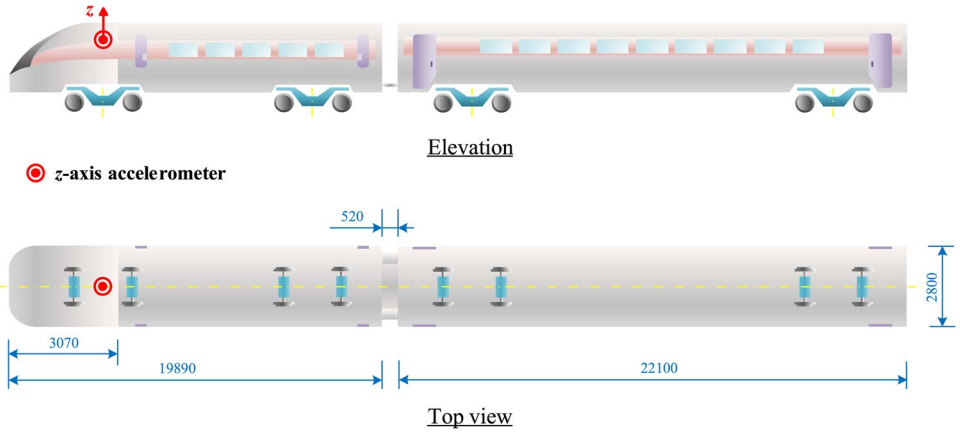
The monitoring setup, as illustrated in Fig. 4a, shows the measurement point located centrally in the driver's cab. Figures 4b and 4c display two uni-axis accelerometers used for vertical acceleration measurement, along with the data acquisition system employed during the experiment. The data collection process strictly adhered to the ISO 2631-1:1997 standard, with acceleration measurements taken from a central position on the cockpit floor. These measurements were used to compute the comfort index factor a_w, which serves as a real-time indicator of passenger ride comfort. To ensure the consistency and reliability of the collected data, we carefully selected the time window of 9:00 a.m. to 10:00 a.m. daily, as this period coincides the peak passenger density and maximum vehicle load, which are critical factors influencing the vehicular ride comfort, according to statistical data provided by the underground train company. This time window was chosen based on empirical observations of the daily passenger flow and vehicle load patterns, ensuring that the data captured during this period would accurately represent the most challenging operational conditions for the vehicle. The data collection spanned ten consecutive days on the conventional TS track, followed by an additional eight days after the installation of the FST track, providing a comprehensive dataset for comparative analysis. Throughout the data collection period, the trains operated at a consistent speed between 70 and 80 km/h to minimize the influence of speed variations on the recorded acceleration data.

This study was focused on the vertical acceleration response, the accelerometers were utilized to capture data at a sampling rate of 20,000 Hz. This high-frequency sampling facilitated the calculation of the acceleration data for the determination of a_w.

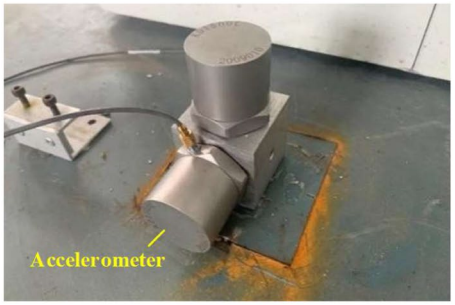
3.2 ISO 2631-1:1997 standard

In the context of underground transportation, the primary focus lies in evaluating the comfort experienced by passengers during their ride. Consequently, the ISO 2631-1:1997 standard was employed to establish methodologies for quantifying the whole-body vibration (WBV) concerning the following aspects: the impact on human's health and comfort, the likelihood of perceiving the vibration, and the occurrence of motion sickness, and outlines the essential elements that collectively determine the acceptable level of vibration exposure.

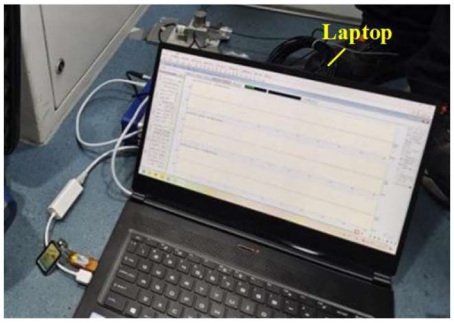
Initially, several parameters were computed using equations 18 to 22 to assess the characteristics of the frequency-weighted accelerations, which included the root means square (RMS) value a_w, the crest factor (X_{max}/X_{rms}), the vibration dose value (VDV) and the maximum transient vibration value (MTVV). The calculations were performed



(a) Configuration of measurement point



(b) Accelerometer



(c) Laptop

Fig. 4 Accelerometers and data acquisition system for online monitoring of an underground train

specifically for accelerations in the z-direction, with a frequency-weighted filter W_k applied.

As previously discussed, the underlying principle of the FST is to reduce the transmission of vibrations from the train to the structure by integrating a rigid spring-oriented perpendicular to the ground. However, it has been observed that although FST can effectively diminish vibrations in the tunnel structure, it can also lead to an increase in vibrations of the train in the direction perpendicular to the ground. Therefore, in light of the vibration characteristics associated with FST, this study primarily concentrates on analyzing vibrations of the train in the direction perpendicular to the ground.

$$a_w = \left[\frac{1}{T}\int_0^T a_w^2(t)dt\right]^{1/2} \tag{18}$$

$$\mathrm{CF} = X_{peak}/X_{rms} \tag{19}$$

$$\mathrm{VDV} = \left[\int_0^T a_w^4(t)dt\right]^{1/4} \tag{20}$$

$$\mathrm{MTVV} = \max\left[a_w(t_0)\right] \tag{21}$$

$$a_w(t_0) = \left[\frac{1}{\tau}\int_{t_0-\tau}^{t_0} a_w^2(t)dt\right]^{1/2} \tag{22}$$

where $a_w(t)$ is the frequency-weighted accelerations with respect to time; $a_w(t_0)$ is the operating RMS values of frequency-weighted accelerations with respect to time; t is the integration time; t_0 is the time of observation or the instantaneous time. The steps of calculating comfort index factor a_w are shown in the Table 3.

In Table 3, The CF represents the ratio of the peak vibration measurement to the RMS vibration measurement, expressed as CF = Peak/RMS. This CF value is utilized to determine the suitable analysis method to be employed.

To identify the most appropriate quantification approach, several parameters were evaluated, including CF, MTVV/a_w, and VDV/($a_w T^{1/4}$). The acceleration data were collected over a period of 18 days, encompassing different railway slab tracks, with one hour of train operation per day. Within each hour, the comfort index factor a_w was calculated at regular intervals of every 3 min. Before analyzing the original time domain vibration data, several preprocessing steps were applied to ensure data quality and consistency. First, noise filtering was employed to eliminate high-frequency noise and external environmental interference. A Butterworth low-pass filter with a cutoff frequency of 80 Hz was applied to

Table 3 Comfort index a_w factor evaluation process

ISO 2631-1:1997	
Step 1	Measure and store accelerations in the z-direction at the seat/operator interface.
Step 2	Calculate the frequency weighted accelerations using the frequency weighting filters.
Step 3	Calculate the RMS value (a_w), the VDV and the MTVV.
Step 4	Analyze values of CF, MTVV/a_w and VDV/($a_w T^{1/4}$) to be determine the suitable quantification method. For a crest factor greater than 9, VDV values should be used. For a CF less than or equal to 9, the RMS method should be used

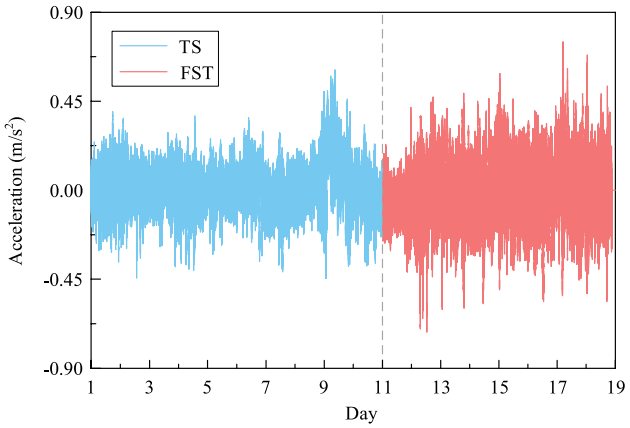


Fig. 5 Original signals of collected acceleration data

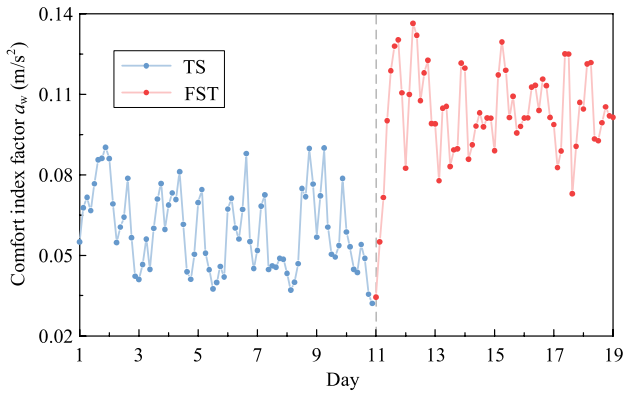


Fig. 6 Comfort index factor a_w of traditional slab and floating slab track

remove high-frequency noise beyond the ISO 2631-1:1997 standard's effective range (0.1–80 Hz). Then, the filtered data was normalized using the Z-score method to ensure that all features, particularly acceleration measurements, were on a comparable scale, facilitating the accurate computation of the comfort index a_w. The processed acceleration signals in the time domain can be observed in Fig. 5, while the evaluation of a_w is depicted in Fig. 6.

According to ISO 2631-1:1997, the RMS approach is considered suitable for evaluating the efforts of WBV if the CF remains below 9. However, the RMS approach might underestimate the impact of WBV when MTVV/a_w exceeds 1.5 or VDV/($a_w T^{1/4}$) exceeds 1.75. As a result of the assessment, it is evident that the CF of the acceleration time-series for TS and FST was 5.46 (<9), as shown in Table 4. Furthermore, the VDV/($a_w T^{1/4}$) factor was 0.5254 (<1.5). The MTVV/a_w factor was 1.491 (<1.75). Therefore, none of these values surpassed the respective predefined limits. Considering these results, it can be reasonably concluded that the selected method remains appropriate for this case.

Table 4 Evaluated factor values of time-series

Whole-body vibration evaluation factor	Value
Crest factor CF	5.46
Composite weighted level a_w (m/s^2)	0.08083
Maximum transient vibration MTVV (m/s^2)	0.1367
Fourth power vibration dose VDV ($m/s^{1.75}$)	0.5254
MTVV/a_w	1.491

Several studies have compiled information regarding the thresholds associated with various levels of predicted "health risk" based on the "health guidance caution zone" values outlined in ISO 2631-1:1997 [47]. These summarized findings can be found in Table 5. It is important for the predicted health risk in the aforementioned study case to be lower than the values specified in Table 5.

4 Analysis results and discussions

The DEWESoft® system was employed to record the vibration signals of the train body using the suggested approach in Sect. 3. The data collection process involved capturing acceleration time-series data of one hour daily from the sensors installed on both TS and FST over a period of 18 days. Specifically, 10 days of data were collected on TS, while 8 days were dedicated to FST.

In accordance with the guidelines specified in the ISO 2631-1:1997 standard, the time-series data that was collected underwent a conversion process to derive the comfort index factor value, denoted as a_w. This computation was carried out at intervals of three minutes, with 50 % overlap

Table 5 Limits associated with different levels of predicted "health risk" according to the ISO 2631-1 (after [47])

ISO 2631-1 assessment of adverse health effects	Terminology used in this paper to describe the predicted health risks	ISO 2631-1	
		a_w index value (m/s^2)	VDV ($m/s^{1.75}$)
"For exposures below the zone (HGCZ), health effects have not been clearly documented and/or objectively observed"	Low	<0.45	<8.5
"... in the zone (HGCZ), caution with respect to potential health risks is indicated"	Moderate	0.45−0.90	8.5−17.0
"... above the zone (HGCZ) health risks are likely"	High	>0.9	>17.0

between consecutive intervals. Consequently, a series of travel quality indices were generated, representing the operational characteristics of the monitored trains across each time period depicted in Fig. 6.

Subsequently, the Bayesian optimization-based BiLSTM technique/framework proposed in this study was utilized to predict future observations, the complete training procedure shown in Fig. 7.

The specific illustrations of each step can be seen as follows:

1. Initially, the acceleration data were collected from a train operating under a TS track and processed to compute the comfort index factor a_w, which served as the training input for a BiLSTM model. Also, the Bayesian parameter optimization method was then employed to identify the optimal hyperparameters for the BiLSTM model, enhancing its predictive accuracy.
2. Subsequently, a distinct BiLSTM model was established and trained exclusively for trains operating on an FST, utilizing fixed hyperparameters determined through the prior optimization process. To enhance the dataset, the comfort index factor a_w values from trains operating under both slab conditions were combined. This augmented dataset was then employed to make predictions using the BiLSTM model, utilizing the already established and fixed hyperparameters.
3. Utilizing the structural characteristics of the LSTM prediction algorithm, sequential predictions were generated by considering the preceding data prior to each subsequent observation. The resulting prediction errors were recorded, and the turning point within these predicted values was identified using the three-sigma theorem. Subsequently, a comparison was conducted between the comfort index factor a_w before and after this turning point.
4. To determine the correlation between the two datasets, Bayesian factors were calculated, and compared with the standard values in Table 2.

To identify the most suitable training ratio for the training set, it is crucial to compare the accuracy of the results obtained from different training ratios. These ratios were selected at intervals of 5 %, ranging from 50 to 90 %. Upon verification, it was observed that the correlation between the data reached its highest point at approximately 80 % of the training ratio. Subsequently, the correlation gradually decreased beyond that point, as depicted in Fig. 8. As a result, a training data ratio of 80 % was utilized during the model establishment process. It is worth mentioning that decline in performance at the 60 % training ratio, can be attributed to the lower quality of the data collected during days 9, 10, and 11. This portion of the dataset exhibited inconsistencies, which negatively impacted the model's ability to learn effectively. However, as the training ratio increases beyond 60 %, more high-quality data is incorporated into the training set, leading to a significant improvement in model accuracy. This demonstrates the

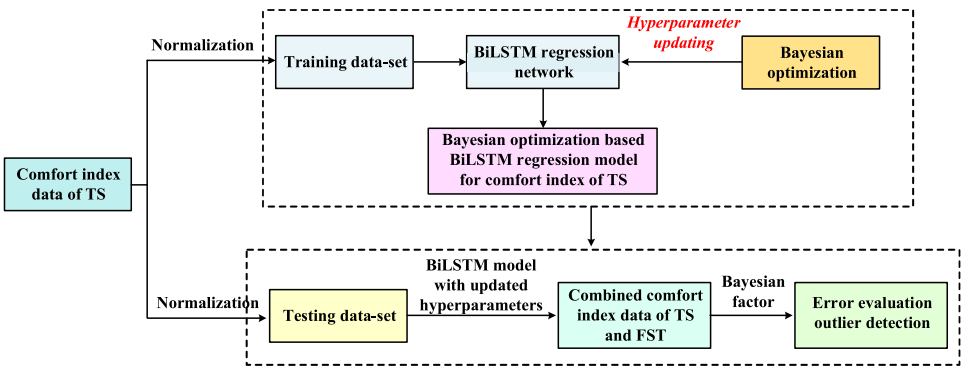


Fig. 7 Bayesian based BiLSTM training process

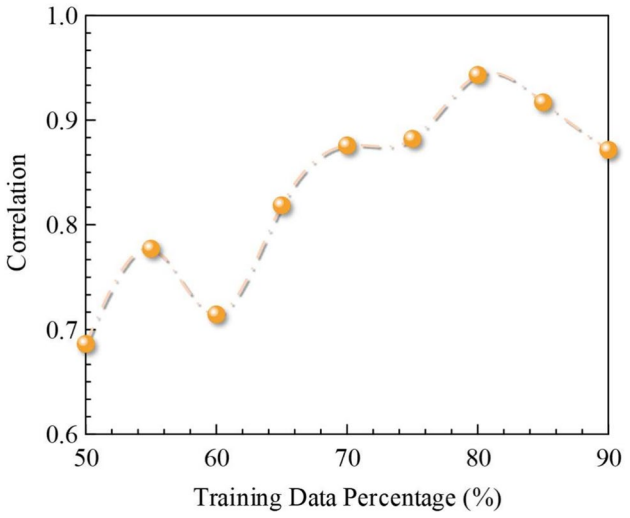


Fig. 8 Evaluation of training data ratio result

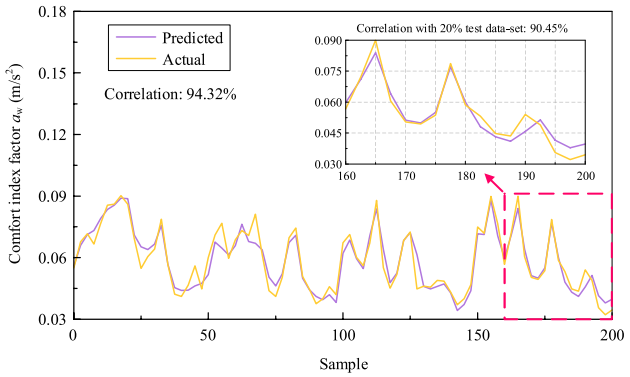


Fig. 9 Rank correlation of TS data

model's ability to generalize and become more resilient to lower-quality data.

After establishing the optimal training ratio, the Bayesian optimization-based BiLSTM method was utilized to conduct prediction and perform a targeted hyperparameter search, focusing specifically on the TS segment. This approach allowed for the identification of the most suitable hyperparameters for the model, ensuring optimal performance.

Figure 9 illustrates the predicted results of the one-step forecasting process alongside the corresponding actual outcomes for the comfort index factor a_w. These values were derived from the original vibration signals at regular time intervals. Subsequently, Bayesian optimization techniques were utilized, resulting in the determination of the optimal values for the hyperparameters, as presented in Table 6.

Following the Bayesian parameter optimization process, the model achieved its highest correlation of 94.32 % by using the aforementioned hyperparameter settings, the 20 % test dataset also showing a correlation of up to 90.45

Table 6 Optimized hyperparameters under Bayesian optimization

Hyperparameters	Range of values	Optimal values
Number of layer(s)	[1, 4]	1
Number of units	[50, 300]	63
Number of BiLSTM layers	[1, 2]	1
Initial learning rate	[10^{-3}, 1]	0.0243158
Regularization	[10^{-10}, 10^{-2}]	1.6720×10^{-10}

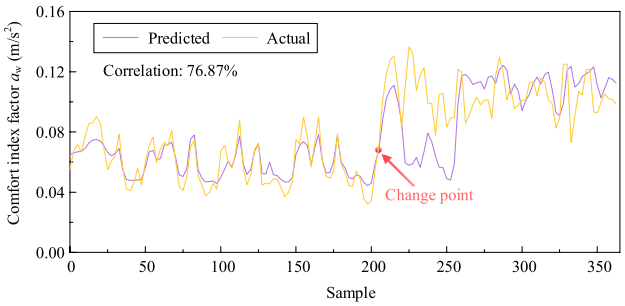


Fig. 10 Rank correlation of data combined TS and FST

%. Then, the optimal hyperparameters were established as the fixed parameters for the combined TS and FST dataset, which remained unchanged throughout the analysis. The range of layers [1,4] balances model complexity and overfitting risks, as deeper networks may fail to converge with limited vibration data (18 days). The unit range [50,300] is empirically set to cover both underfitting and overfitting scenarios, determined by sampling across the range using a Gaussian process model. For number of BiLSTM layers, the optimization indicated that a single BiLSTM layer was sufficient to capture bidirectional dependencies, with additional layers not providing a significant performance gain relative to increased computational complexity. Then the model is optimized through Bayesian sampling by iteratively refining the learning rate to achieve fast, stable convergence. The selected value minimized the loss function as measured on the validation set across the iterations. A lower regularization value was preferred as it resulted in a lower loss while preventing overfitting. The process balanced the need for weight penalization with the capacity required to adequately learn from the data.

Upon incorporating the FST data into the entire dataset while maintaining the fixed hyperparameters, the correlation decreased from 94.32 to 76.87 %, which is shown in the Fig. 10. The observed decrease in performance suggests that the inclusion of FST data in the overall dataset has a substantial impact on the prediction accuracy of the model. This decline is not an unintentional setback but a deliberate strategy employed to identify change points in operational conditions. By examining the influence of introducing data

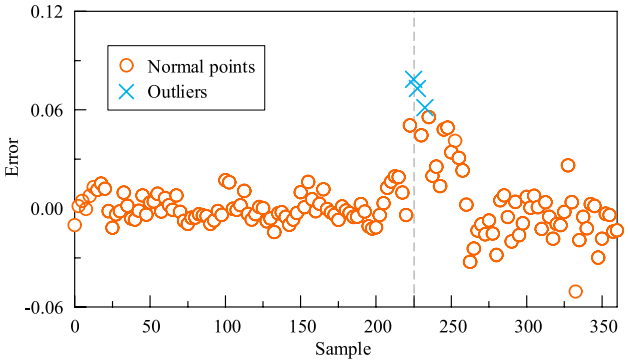


Fig. 11 Error evaluation of data combined TS and FST

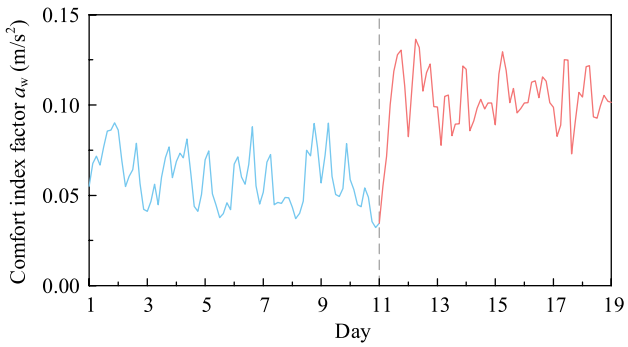


Fig. 12 a_w index value evaluation of combined TS and FST

from the complex track system on the model's performance, which had previously been calibrated for different conditions, a calculated approach was taken to assess the impact on prediction accuracy. While these solid results demonstrate the model's capability, they also reveal potential limitations. The model could be prone to overfitting when trained on a consistent dataset, and may underfit if the data does not capture sufficient operational variability. Future work will benefit from incorporating additional cross-validation, data augmentation, and regularization strategies to mitigate these issues and improve generalization.

The analysis conducted in Fig. 11 yielded a standard deviation of 0.0153 for the comfort index factor a_w within the TS part. The evaluation error reached its peak at the 225^{th} sample. By employing the three-sigma control limits, the turning point was identified as the moment when the error exceeded three times the standard deviation. Importantly, these turning points correspond closely with the actual observed changes in the operating conditions.

Figure 12 illustrates the turning point that divides the entire dataset into two distinct parts. To assess the correlation between the data before and after the turning point, it is necessary to calculate the Bayesian factors of these two datasets, using equation 16, which yields a Bayesian factor $b_{10} = 4.92$. It falls within the range [3, 10], indicating robust evidence of a "change" occurring between the two parts, which further validates the reliability of the aforementioned turning point.

5 Conclusions

This study introduces an innovative framework that employs a Bayesian-optimized BiLSTM algorithm for the monitoring of subway operational conditions and precise assessment of vibration comfort levels.

By integrating real-time data collected via installed sensing devices, the study has facilitated the extraction of key operational indicators that adhere to the ISO 2631-1:1997 standard for comfort quality.

The BiLSTM model, at the core of this framework, has proven to be essential in enhancing the precision of the analysis, owing to its fine-tuned hyperparameters achieved through Bayesian optimization. The model's ability to facilitate predictive analysis and real-time monitoring in condition forecasting and operational efficiency. Quantitative analysis indicates that the optimized bi-directional LSTM model attains a correlation of up to 94.32 % across the entire dataset and 90.45 % on the test dataset.

The model's effectiveness in identifying transitional points in train operations and consistently differentiating slab conditions has been proven through validation. The proposed methodology integrates vibration data and a ride comfort index to precisely detect changes in the operational condition of the train, enabling early detection of potential faults.

The findings underscore the effectiveness of this approach in accurately determining the train's operational status. This research highlights the potential of machine learning algorithms to enhance the safety and efficiency of underground railways and emphasizes the critical role of continuous monitoring and predictive maintenance in transportation.

Acknowledgements This study was funded by the Start-up Fund for RAPs under the Strategic Hiring Scheme of The Hong Kong Polytechnic University (grant number 1-BD22). The authors also acknowledge financial support from the Innovation and Technology Commission (ITC) of the Hong Kong SAR Government to the Hong Kong Branch of the Chinese National Rail Transit Electrification and Automation Engineering Technology Research Center (grant number K-BBY1).

Funding Open access funding provided by The Hong Kong Polytechnic University. This study was funded by the Start-up Fund for RAPs under the Strategic Hiring Scheme of The Hong Kong Polytechnic University (Grant number 1-BD22).

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

1. Barke D, Chiu WK (2005) Structural health monitoring in the railway industry: a review. Struct Health Monit 4(1):81–93
2. Remennikov AM, Kaewunruen S (2008) A review of loading conditions for railway track structures due to train and track vertical interaction. Struct Control Health Monit 15(2):207–234
3. Zhang D, Guo ZH, Ni YQ, Chen ZW, Ao WK, Bordbar A, Zhou FR (2023) Correlation between cargo properties and train overturning safety for a high-speed freight train under strong winds. Eng Appl Comput Fluid Mech 17(1):2221308
4. Salvador P, Naranjo V, Insa R, Teixeira P (2016) Axlebox accelerations: Their acquisition and time-frequency characterisation for railway track monitoring purposes. Measurement 82:301–312
5. Zhang L-H, Wang Y-W, Ni Y-Q, Lai S-K (2018) Online condition assessment of high-speed trains based on bayesian forecasting approach and time series analysis. Smart Struct Syst 21(5):705–713
6. Meymand SZ, Keylin A, Ahmadian M (2016) A survey of wheel-rail contact models for rail vehicles. Veh Syst Dyn 54(3):386–428
7. Azimi M, Eslamlou AD, Pekcan G (2020) Data-driven structural health monitoring and damage detection through deep learning: State-of-the-art review. Sensors 20(10):2778
8. Wang Y, Ni Y (2020) Bayesian dynamic forecasting of structural strain response using structural health monitoring data. Struct Control Health Monit 27(8):e2575
9. Xia Q, Wu W-L, Li F-N, Xia Y, Ding X-L, Lam WH, Chung W-H, Xu Y-L (2021) System design and demonstration of performance monitoring of a butterfly-shaped arch footbridge. Struct Control Health Monit 28(7):e2738
10. Finotti RP, Cury AAC, Barbosa FdS (2019) An shm approach using machine learning and statistical indicators extracted from raw dynamic measurements. Latin Am J Solids Struct 16:e165
11. Meixedo A, Ribeiro D, Santos J, Calçada R, Todd M (2022) Structural health monitoring strategy for damage detection in railway bridges using traffic induced dynamic responses 389–408
12. Ao WK, Hester D, O'Higgins C, Brownjohn J (2024) Tracking long-term modal behaviour of a footbridge and identifying potential shm approaches. J Civ Struct Heal Monit 14(5):1311–1337
13. He S, Ao WKA, Ni Y-Q A unified label noise-tolerant framework of deep learning-based fault diagnosis via a bounded neural network, IEEE Transact Instrum Meas
14. O'Higgins C, Hester D, Ao WK, McGetrick P (2024) A method to maximise the information obtained from low signal-to-noise acceleration data by optimising ssi-cov input parameters. J Sound Vib 571:118101
15. O'Higgins C, Hester D, McGetrick P, Cross EJ, Ao WK, Brownjohn J (2023) Minimal information data-modelling (mid) and an easily implementable low-cost shm system for use on a short-span bridge. Sensors (Basel, Switzerland) 23(14):6328
16. O'Higgins C, Hester D, McGetrick P, Ao WK, Cross EJ (2024) Refinement and validation of the minimal information data-modelling (MID) method for bridge management. Sensors 24(12):3879.
17. Ao WK, Pavic A, Kurent B, Perez F (2023) Novel frf-based fast modal testing of multi-storey clt building in operation using wirelessly synchronised data loggers. J Sound Vib 548:117551
18. Manthey M, Flamand O, Jalil A, Pavic A, Ao WK (2021) Effect of non-structural components on natural frequency and damping of tall timber building under wind loading, World Conference on Timber Engineering
19. Ao WK, Reynolds P (2017) Analytical and experimental study of eddy current damper for vibration suppression in a footbridge structure. In: Dynamics of Civil Structures, Volume 2: Proceedings of the 35th IMAC, A Conference and Exposition on Structural Dynamics 2017, pp 131–138
20. Ao WK, Tang Q-C, Pavic A (2024) Decentralised H robust control of MTMDs for mitigating vibration of a slender MDOF floor configuration. Thin-Walled Struct 203:112226
21. O'Higgins C, Hester D, Ao WK, McGetrick P, Robinson D (2019) Inherent uncertainty in the extraction of frequencies from time-domain signals. Infrastruct Asset Manage 8(3):121–32
22. Hewamalage H, Bergmeir C, Bandara K (2021) Recurrent neural networks for time series forecasting: Current status and future directions. Int J Forecast 37(1):388–427
23. Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci 79(8):2554–2558
24. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780
25. Qin C, Chen L, Cai Z, Liu M, Jin L (2023) Long short-term memory with activation on gradient. Neural Netw 164:135–145
26. Silka J, Wieczorek M, Wozniak M (2022) Recurrent neural network model for high-speed train vibration prediction from time series. Neural Comput Appl 34(16):13305–13318
27. Sharma S, Sen S (2023) Real-time structural damage assessment using lstm networks: regression and classification approaches. Neural Comput Appl 35(1):557–572
28. Wang C, Ansari F, Wu B, Li S, Morgese M, Zhou J (2022) Lstm approach for condition assessment of suspension bridges based on time-series deflection and temperature data. Adv Struct Eng 25(16):3450–3463
29. Lim J-Y, Kim S, Kim H-K, Kim Y-K (2022) Long short-term memory (lstm)-based wind speed prediction during a typhoon for bridge traffic control. J Wind Eng Ind Aerodyn 220:104788
30. Guo A, Jiang A, Lin J, Li X (2020) Data mining algorithms for bridge health monitoring: Kohonen clustering and lstm prediction approaches. J Supercomput 76(2):932–947
31. Luo H, Huang M, Zhou Z (2019) A dual-tree complex wavelet enhanced convolutional lstm neural network for structural health monitoring of automotive suspension. Measurement 137:14–10327
32. Tien T, Quang T, Ngoc L, Ngoc H (2024) Time series data recovery in shm of large-scale bridges: Leveraging gan and bi-lstm networks. Structures 63:106368
33. Ghazimoghadam S, Hosseinzadeh SAA (2024) A novel unsupervised deep learning approach for vibration-based damage diagnosis using a multi-head self-attention lstm autoencoder. Measurement 229:114410

34. Luo H, Huang M, Zhou Z (2018) Integration of multi-gaussian fitting and lstm neural networks for health monitoring of an automotive suspension component. J Sound Vib 428:87–103
35. International organization for standardization, mechanical vibration and shock-evaluation of human exposure to whole body vibration-part 1: general requirements, iso 2631-1:1997
36. Salman AG, Heryadi Y, Abdurahman E, Suparta W (2018) Single layer & multi-layer long short-term memory (lstm) model with intermediate variables for weather forecasting. Proced Comput Sci 135:89–98
37. Brownlee J (2018) Deep learning for time series forecasting: predict the future with MLPs, CNNs and LSTMs in Python
38. Yildirim O (2018) A novel wavelet sequence based on deep bidirectional lstm network model for ecg signal classification. Comput Biol Med 96:189–202
39. Graves A, Jaitly N, Mohamed A-r (2013) Hybrid speech recognition with deep bidirectional lstm 273–278
40. Wu J, Chen X-Y, Zhang H, Xiong L-D, Lei H, Deng S-H (2019) Hyperparameter optimization for machine learning models based on bayesian optimization. J Electron Sci Technol 17(1):26–40
41. Astudillo R, Frazier P (2019) Bayesian optimization of composite functions 354–363
42. Bergstra J, Bardenet R, Bengio Y, Kégl B Algorithms for hyper-parameter optimization. Adv Neural Inform Process Syst 24
43. Lodewyckx T, Kim W, Lee MD, Tuerlinckx F, Kuppens P, Wagenmakers E-J (2011) A tutorial on bayes factor estimation with the product space method. J Math Psychol 55(5):331–347
44. Ly A, Verhagen J, Wagenmakers E-J (2016) Harold jeffreys's default bayes factor hypothesis tests: Explanation, extension, and application in psychology. J Math Psychol 72:19–32
45. Ly A, Wagenmakers E-J (2022) Bayes factors for peri-null hypotheses. TEST 31(4):1121–1142
46. Shewhart WA, Deming WE (1986) Statistical method from the viewpoint of quality control
47. Eger T, Stevenson J, Boileau PE, Salmoni A (2008) Predictions of health risks associated with the operation of load-haul-dump mining vehicles: Part 1-analysis of whole-body vibration exposure using iso 2631–1 and iso-2631-5 standards. Int J Ind Ergon 38(9–10):726–738