IEEE Access
Multidisciplinary ⁝ Rapid Review ⁝ Open Access Journal

Received November 14, 2019, accepted December 8, 2019, date of publication December 18, 2019, date of current version December 31, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2960537

Deep Learning for Track Quality Evaluation of High-Speed Railway Based on Vehicle-Body Vibration Prediction

SHUAI MA[1], LIANG GAO[1], XIUBO LIU[2], AND JING LIN[3], (Member, IEEE)

[1]School of Civil Engineering, Beijing Jiaotong University, Beijing 100044, China
[2]Infrastructure Inspection Research Institute, China Academy of Railway Sciences Corporation Limited, Beijing 100081, China
[3]Department of Civil, Environmental and Natural Resources Engineering, Luleå University of Technology, 97187 Luleå, Sweden

Corresponding author: Shuai Ma (ms5274@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant U1734206 and Grant 51827813, and in part by the Fundamental Research Funds for the Central Universities under Grant 2018JBM042.

ABSTRACT Track quality evaluation is fundamental for track maintenance. Around the world, track geometry standards are established to evaluate track quality. However, these standards may not be capable of detecting some abnormal track geometry conditions that can cause considerable vehicle-body vibration. And people gradually realized that track quality evaluation should be based not only on track geometry but also on vehicle performance. Vehicle-body vibration prediction is beneficial for locating potential track geometry defects, and the predicted accelerations can be used as an auxiliary index for assessing track quality. For this purpose, this paper gives a method to predict vehicle-body vibration based on deep learning, which represents one of the newest areas in artificial intelligence. By integrating convolutional neural network (CNN) and long short-term memory (LSTM), a CNN-LSTM model is proposed to make accurate and point-wise prediction. To achieve the optimal performance and explore the internal mechanism of the model, structural configurations and inner states are extensively studied. CNN-LSTM can take advantage of the powerful feature extraction capacity of CNN and LSTM, and outperforms the fully-connected neural network and the plain LSTM on the experimental data of a high-speed railway. In detail, CNN-LSTM has superior performance in predicting vertical vehicle-body vibration below 10 Hz and lateral vehicle-body vibration below 1 Hz. Moreover, analysis shows that the predicted vehicle-body acceleration can act as a performance-based evaluation index of track quality.

INDEX TERMS Track quality evaluation, track geometry, vehicle-body vibration, convolutional neural network (CNN), long short-term memory (LSTM), CNN-LSTM.

I. INTRODUCTION

With the development of high-speed railway (HSR), people are paying more attention to the safety and ride comfort. As railway track provides the running surface for the vehicle, track geometry is the main cause of vehicle responses, among which vehicle-body vibration is the evaluation index for ride comfort. Nowadays, railway administrations utilize comprehensive inspection train (CIT) to routinely inspect track geometry, such as longitudinal level, alignment, gage, cross-level, and so on. Statistical quantities of the inspection data are then calculated to assess track quality and guide maintenance according to predefined limits in current track geometry standards [1]–[3]. However, these standards may be incapable of identifying some potentially adverse track geometry conditions that can cause undesirable vehicle response. Because vehicle response is simultaneously affected by several nonlinear excitations and various frequency components of track geometry [4], [5], simple track geometric statistics are not always an accurate indicator for vehicle performance. This situation is a hindrance to fulfilling performance-based evaluation of HSR track. Since the last decade, there has been a growing trend in track quality evaluation based not only on track geometry but also on vehicle response. And this necessitates a prediction model that can effectively relate track geometry to vehicle-body vibration, as ride comfort is one of the most concerned issues for HSR. There are two benefits for this task. The first is identifying latent locations where

The associate editor coordinating the review of this manuscript and approving it for publication was Faisal Khan.

significant vehicle-body acceleration (VBA) frequently occurs. The second is utilizing the predicted accelerations as an additional index for performance-based track quality evaluation. Furthermore, the model is valuable in understanding the complex correlation between track geometry and vehicle-body vibration, and in further studying the contributions of different track geometry parameters to vehicle-body vibration.

In the literature, generally two kinds of methodologies, based on mechanism model and data model respectively, are proposed to predict vehicle response by using track geometry. Mechanism models simulate the vehicle-track system with motion equations and typically use numerical iteration methods to calculate vehicle response [6]. With the support of modern simulation packages, such as VAMPIRE and SIMPACK [7], it is more convenient to simulate various vehicle types and speed conditions. However, the performance of mechanism models depends largely on the authenticity and reliability of system parameters. Besides, these models tend to be time consuming and therefore mostly used for off-line analysis. Data models can be further categorized into signal processing-based models and machine learning-based models. The former simulates the vehicle-track dynamic characteristics by using transfer functions, which can be obtained either in the frequency domain by using FFT or in the time domain by constructing a parametric filter function based on system identification [8]. Correspondingly, the prediction process can be carried out either by using IFFT or by the filter function. Signal processing-based models share the merits of calculation efficiency and conceptual clarity, but they are inherently linear models and only applicable to constant speed conditions. The latter takes advantage of intelligent models, such as artificial neural network and support vector machine, to recognize complex patterns and nonlinear relationships between track geometry and vehicle response [4], [9]. Among them, fully-connected neural network (FNN) has been acting as a core role in the well-known performance-based track geometry (PBTG) technology [4] for more than ten years. Nonetheless, PBTG is based on statistical indices of track geometry, such as the maximum and 95th percentile, which are typically generated from a railway section, thus unable to make point-wise prediction yet. Besides, machine learning-based models tend to take in well-crafted data features, which may require time consuming calculations or profound expertise. To make the best of the above two methodologies, some researchers also propose to combine mechanism model with system identification [10]–[12]. But these endeavors are focused on predicting wheel-rail forces only and have not been extended to vehicle-body accelerations yet.

The last decade has witnessed revolutionary advances in deep learning, which represents one of the newest trends in machine learning and artificial intelligence. Every now and then, new deep learning models are being born, outperforming state-of-the-art techniques in the fields of computer vision and natural language processing [13]. Convolutional neural network (CNN) and long short-term memory (LSTM) are classic deep learning models particularly designed for image and sequential data respectively, and have already been adopted in safety surveillance [14], disease prediction [15], human kinematics interpretation [16], health condition monitoring [17], and vehicle fault diagnosis [18], [19]. These models tend to demonstrate better performance than traditional machine learning models, for example, some deep convolutional nets have even surpassed human-level performance in visual recognition tasks [20]. This drops a hint for the possibility of relating track geometry to vehicle-body vibration with the support of deep learning models. The rationales lie in that vehicle-body vibration can be collected as spatially sequential data and has correlations with track geometry waveforms. Fortunately, LSTM and CNN are specialized in learning sequential information and shape characterizations respectively. Nonetheless, few research has been found that focused on vehicle-body vibration prediction with deep learning models.

Considering the above motivations, we extensively studied CNN and LSTM for predicting vehicle-body vibration by using track geometry. The main contributions of this paper include:

1. An effective model combining CNN and LSTM is proposed to predict VBA. In order to obtain the optimal performance, some structural hyper-parameters are carefully studied. By fine-tuning, this model can also be utilized in predicting other vehicle responses, such as wheel-rail forces and so on.

2. The proposed model outperforms FNN of PBTG and plain LSTM in terms of accuracy. Moreover, the model allows point-wise prediction, thus facilitates precise track maintenance.

3. By utilizing the predicted VBA as an evaluation index, the correlation between evaluation quantities and vehicle response is enhanced, thus the predicted VBA can be taken as a performance-based index of track quality.

The rest of the paper is organized as follows: In Section II, typical deep learning models and the proposed model are introduced. In Section III, the specifics of the data source, the process of configuring model hyper-parameters, as well as model assessment are given. Besides, visualization of model inner states and comparison between track geometry-based and VBA-based evaluation indexes are illustrated. Finally, the conclusions are given in Section IV.

II. THE PROPOSED MODEL

A. CNN FOR SHAPE FEATURE LEARNING

CNN is specifically designed to deal with the variability of two-dimensional (2D) shapes, and typically has a standard structure – stacked convolutional layers and pooling layers followed by several fully-connected layers, as shown in Figure 1. Convolutional layers act as the feature extractor, which transforms the inputs into multi-channel activations containing distinctive features. There are several kernels in each convolutional layer, and each kernel contains a set of

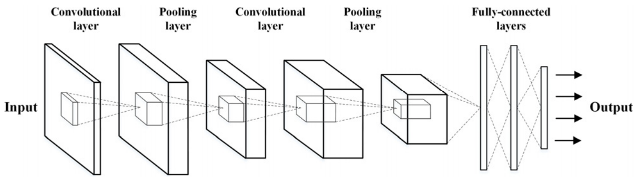

FIGURE 1. Typical architecture of CNN for 2D data.

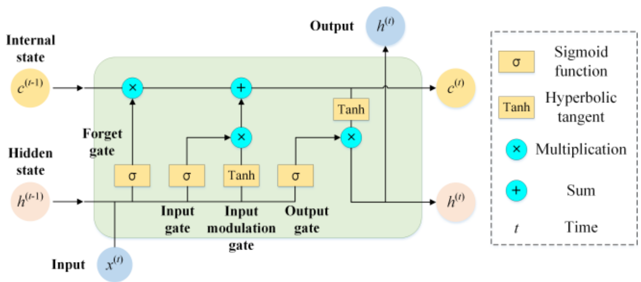

FIGURE 2. Classical architecture of a LSTM cell.

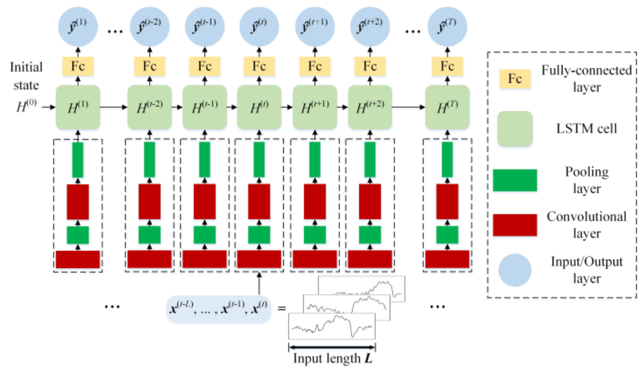

FIGURE 3. Schematic diagram of CNN-LSTM model.

learnable weights. When applied to sequential data, CNN conducts 1D convolution with one-dimensional kernels [21]. Following convolutional layers, pooling layers reduce the resolution of activations for the purpose of detecting higher-scale features. Finally, fully-connected layers are used as classifiers, which synthesize a complex decision surface to categorize high-dimensional features.

With the property of shift and scale invariance [22], CNN is able to extract geometric shape features. The features are contained in local track geometry waveforms and have correlations with vehicle-body vibration. For HSR vehicles, vehicle-body vibration tends to be sensitive to multi-wavelength components of track geometry, thus necessitating stacking convolutional layers for a balanced extraction of multi-scale shape features. For the optimal depth of stacking layers, it can be determined by using the trial method and will be discussed in Section III.

B. LSTM FOR SEQUENTIAL FEATURE LEARNING

Compared with feed-forward neural networks (like FNN in PBTG), recurrent neural network (RNN) is strengthened by a time step edge that introduces a notion of time. In order to jump over the optimization hurdles that plague RNN, long short-term memory (LSTM) architecture was formulated [23]. The central idea behind the LSTM architecture is a memory cell maintaining its state over time and nonlinear gating units which regulate the information flow into and out of the cell, as shown in Figure 2.

The information flow inside a LSTM cell can be described as follows:

$$z^{(t)} = \mathrm{Tanh}(W_{zh}h^{(t-1)} + W_{zx}x^{(t)} + b_z) \tag{1}$$
$$i^{(t)} = \sigma(W_{ih}h^{(t-1)} + W_{ix}x^{(t)} + b_i) \tag{2}$$
$$f^{(t)} = \sigma(W_{fh}h^{(t-1)} + W_{fx}x^{(t)} + b_f) \tag{3}$$
$$o^{(t)} = \sigma(W_{oh}h^{(t-1)} + W_{ox}x^{(t)} + b_o) \tag{4}$$
$$c^{(t)} = f^{(t)} \otimes c^{(t-1)} + i^{(t)} \otimes z^{(t)} \tag{5}$$
$$h^{(t)} = o^{(t)} \otimes \mathrm{Tanh}(c^{(t)}) \tag{6}$$

where W_{zh}, W_{zx}, W_{ih}, W_{ix}, W_{fh}, W_{fx}, W_{oh}, and W_{ox} are weights; b_z, b_i, b_f, b_o are biases; $\mathrm{Tanh}(\cdot)$ and $\sigma(\cdot)$ represent hyperbolic tangent function and sigmoid function. The input modulation gate, input gate, forget gate, and output gate take in input $x^{(t)}$ and hidden state $h^{(t-1)}$ of the previous time, and output $z^{(t)}$, $i^{(t)}$, $f^{(t)}$, and $o^{(t)}$ respectively, which are then synthesized into hidden state $h^{(t)}$ and cell state $c^{(t)}$. With the assistance of these gates, the LSTM cell keeps absorbing new key information and discarding obsolete irrelevant information. Besides, as LSTM maintains its cell state through time, the notorious problem of vanishing/exploding gradient [24] is well settled. This structural design also makes LSTM effective at capturing long-term temporal dependencies.

Inspection data of track geometry and VBA are multi-channel sequences with a fixed spatial interval between every two data points. Each sequence is self-correlated in nature and the VBA at each point can be considered as a posterior probability given the preceding VBAs and track geometry, thus necessitating the LSTM module for sequential feature learning. As stacking multiple LSTM layers enables the learning of higher level sequential features, the optimal depth will be discussed in Section III.

C. CNN-LSTM MODEL

Combining CNN and LSTM, a CNN-LSTM model is proposed for the point-wise prediction of VBA by using track geometry. Suppose x and y as track geometry parameters and vehicle-body accelerations, respectively. The architecture of CNN-LSTM is plotted in Figure 3.

While CNN-LSTM makes point-wise prediction, it inputs a section of track geometry, that is, for every point t, L data points that locate ahead of it together with $x^{(t)}$, are considered. This input mode enables CNN to extract shape features in track geometry waveforms. Average longitudinal level, average alignment, and cross-level constitute the three input channels, which has been validated to have the best performance among all possible combinations. The vertical and lateral vehicle-body acceleration (VVBA and LVBA) constitute the two-channel output for point t.

CNN-LSTM adopts the structure of alternating convolutional and pooling layers. In convolutional layers, 1D kernels with size 5 and stride 1, as well as zero padding, are employed to fulfill the convolution operation. The output of convolutional layers is activated by Rectified Linear Units (ReLUs) [25] to add nonlinearity. Then, the resolution of the activations is halved by a max-pooling layer. Accordingly, the width of the activations will get small as the depth gets large. To compensate for this width reduction, the filter sizes are doubled after each convolution operation, starting from four for the bottom convolutional layer. The activations of the topmost pooling layer are put into stacked LSTM layers, in which the size of hidden state $H^{(t)}$ (including $h^{(t)}$ and $c^{(t)}$) is 128 and $H^{(0)}$ is initialized by zero. Each LSTM cell corresponds to an individual pair of VBAs. Above LSTM layers, two fully-connected layers are taken as a classifier, which finally outputs the predicted VBA $\hat{y}^{(t)}$.

In the training process, dropout [26] randomly blocks the connections between interlayer nodes. This trick is employed in LSTM layers and fully-connected layers to avoid overfitting. Mean square error of predicted VBA $\hat{y}$ and true VBA y, adding the L_2-norm regularization of model parameters, is taken as the loss function:

$$Loss = \frac{1}{T}\sum\nolimits_{t=1}^{T}(y^{(t)} - \hat{y}^{(t)})^2 + \lambda \left\|W\right\|_2^2 \tag{7}$$

where W represents the collection of all the trainable parameters; λ denotes the regularization parameter. The initial values of W are drawn from a normal distribution with a standard deviation of 0.1. Adam optimization method [27] is adopted to train CNN-LSTM and the learning rate is set to 0.001.

III. CASE STUDY

A. DATA SOURCE

Track geometry is one of the most important excitations for HSR vehicles. Track geometry inspection and track quality evaluation are essential for railway administrations to control safety and ensure ride comfort for passengers. Therefore, track geometry, as well as VVBA and LVBA, is routinely inspected by the CIT. Nowadays, more than ten CITs are at service in China. Figure 4 shows a CIT at service and an intercepted section of track inspection data.

The track inspection data of a HSR line in China is selected as the data source. The data volume is about 6 million (300 km), of which ninety percent are for training and the rest are for testing. The inspection data are collected at 0.25 m increments along the track with a relatively stable speed of 300 km/h.

B. MODEL CONFIGURATIONS

Several hyper-parameters that determine the configurations of CNN-LSTM, including the input length L, depth of CNN, depth of LSTM, and the convolutional kernel size, have an impact on model performance. Trial method is used for the coarse range selection and careful determination of these parameters. When studying one parameter, all the other parameters are fixed in a proper range. Four assessment indices for model performance are employed, including Mean Absolute Error (*MAE*), Root Mean Square Error (*RMSE*), Theil Inequality Coefficient (*TIC*), and Pearson Correlation Coefficient (*PCC*). Their definitions are as follows:

$$MAE = \frac{1}{T}\sum\nolimits_{t=1}^{T}\left\|y^{(t)} - \hat{y}^{(t)}\right\| \tag{8}$$

$$RMSE = \sqrt{\frac{1}{T}\sum\nolimits_{t=1}^{T}(y^{(t)} - \hat{y}^{(t)})^2} \tag{9}$$

$$TIC = \frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(y^{(t)} - \hat{y}^{(t)})^2}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(y^{(t)})^2} + \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{y}^{(t)})^2}} \tag{10}$$

$$PCC = \frac{Cov(y, \hat{y})}{\sqrt{Var(y)}\sqrt{Var(\hat{y})}} \tag{11}$$

where y and $\hat{y}$ represent the actual and the predicted VBA, respectively; T is the volume of y; $Cov(\cdot)$ is covariance; $Var(\cdot)$ is variance. For absolute indices, such as *RMSE* and

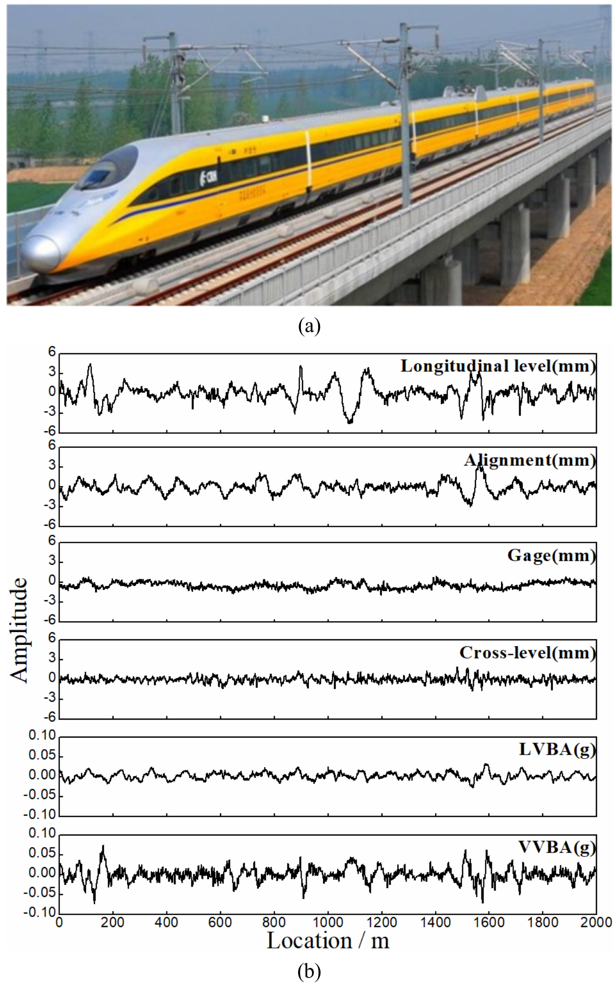

FIGURE 4. Track inspection: (a) comprehensive inspection train; (b) track inspection data.

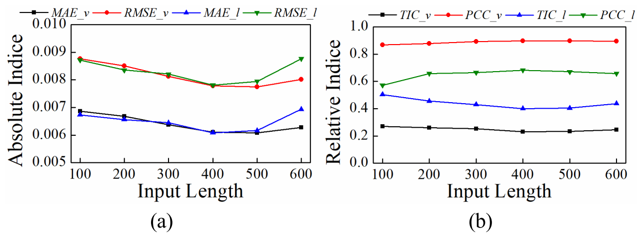

FIGURE 5. Effect of input length on model performance. (a) Absolute indices. (b) Relative indices.

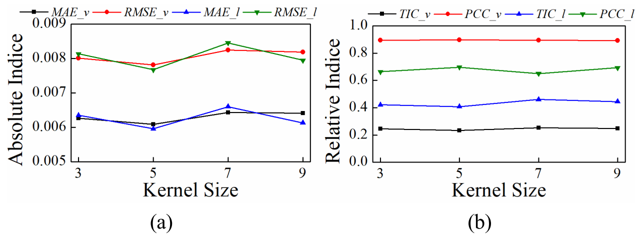

FIGURE 6. Effect of convolutional kernel size on model performance. (a) Absolute indices. (b) Relative indices.

MAE, smaller values correspond to better performance. For relative indices, smaller *TIC* (ranges from 0 to 1) means higher accuracy, and it is the opposite for *PCC* (ranges from -1 to 1).

1) INPUT LENGTH

Several input lengths, including 100, 200, 300, 400, 500, and 600 points, are compared and the assessment indices are given in Figure 5. Abbreviations are employed for simplicity, for example, *MAE_v* and *MAE_l* represent the *MAE* of VVBA and LVBA respectively. It can be seen that CNN-LSTM reaches the best performance when L ranges from 400 to 500. As the upper bound of track geometry inspection wavelength is 120 m, CNN-LSTM employs an input length of 480 points.

2) CONVOLUTIONAL KERNEL SIZE

Different kernel sizes, i.e., 3, 5, 7, and 9, are compared to analyze their influence on model performance, as shown in Figure 6. As standard 1×1 convolution is equivalent to cross channel pooling [28], kernel size 1 is neglected in this study. It can be seen that CNN-LSTM performs the best when the kernel size is 5 (1×5 to fit for 1D convolution). While small kernels are welcomed by newly proposed CNN structures [29]-[30], they may not the best choice for the application of vehicle-body vibration prediction here.

3) STACKING LAYERS OF CNN

There is a trend for CNN that newly proposed structures are getting extremely deep, for example, the deep residual network has more than 1000 stacking layers [31]. Although deep networks are proven to be generally accurate in image recognition, they may be too complicated and thus unsuitable for the application concerned. To determine the optimal depth of CNN, several options ranging from 1 to 5 are compared, as shown in Figure 7.

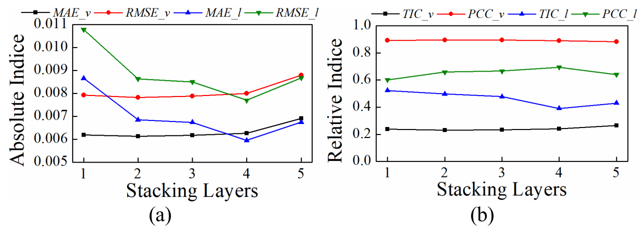

FIGURE 7. Effect of CNN stacking layers on model performance. (a) Absolute indices. (b) Relative indices.

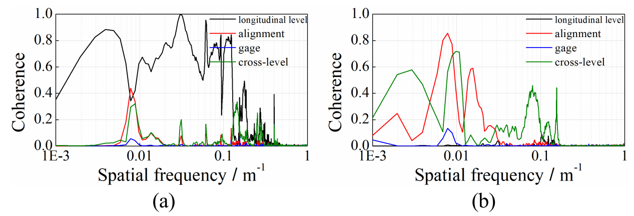

FIGURE 8. Coherence between track geometry and VBA. (a) VVBA. (b) LVBA.

The optimal CNN depth for VVBA and LVBA is 2 and 4 respectively. To explain the difference, coherence between track geometry and VBA is calculated, as given in Figure 8. Coherence between two signals X and Y can be obtained according to the following equations:

$$C_{XY}(\omega) = \frac{|S_{XY}(\omega)|^2}{S_{XX}(\omega)S_{YY}(\omega)} \tag{12}$$

where C_{XY} is coherence; S_{XX} and S_{XY} are the power spectral density of X and Y; S_{XY} is the cross-power spectral density between X and Y; ω is frequency.

According to Figure 8, VVBA is most correlated with longitudinal level in a wide frequency bandwidth, while LVBA is most correlated with alignment and cross-level only in a narrow bandwidth at relatively low frequencies. In the spatial frequency domain, lower frequency correspond to larger wavelength. Besides, it is known that activations in deeper convolutional layers tend to contain more abstract information, which is equivalent to long wavelength contents of track geometry. Accordingly, the activations of the second CNN layer tend to evenly contain multi-wavelength information, while those of the fourth CNN layer tend to have more long-wavelength information. This resonates with the coherence between track geometry and VBA, and may justify the difference in optimal CNN depths for predicting VVBA and LVBA. Considering the accuracy and efficiency of CNN-LSTM, two convolutional layers are employed.

4) STACKING LAYERS OF LSTM

As stacking multiple LSTM layers has the benefit of learning higher-level sequential features, the optimal depth of the LSTM module is investigated. By comparing different depths ranging from 1 to 5, the assessment indices of CNN-LSTM are obtained, as shown in Figure 9. It can be seen that CNN-LSTM has the best performance when stacking two LSTM layers.

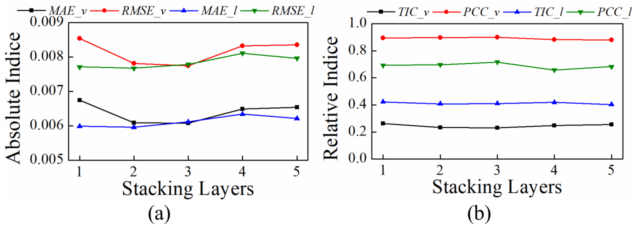

FIGURE 9. Effect of LSTM stacking layers on model performance. (a) Absolute indices. (b) Relative indices.

TABLE 1. Specifics of the CNN-LSTM model.

Layer	Size of layer output (width×channel)	Details
Input	480×3	480 continuous data points of average longitudinal level, average alignment, and cross-level
Convolutional	480×4	Kernel size 1×5, stride 1, ReLU
Max-pooling	240×4	Kernel size 1×2, stride 2
Convolutional	240×8	Kernel size 1×5, stride 1, ReLU
Max-pooling	120×8	Kernel size 1×2, stride 2
Dimension transformation	960×1	Reshape multi-channel activations
LSTM	128×1	Two identical stacking layers, dropout rate 0.2
Fully-connected	128×1	Dropout rate 0.2, ReLU activation
Fully-connected	2×1	Identity activation
Output	2×1	VVBA and LVBA

TABLE 2. Assessment indices of FNN, LSTM, and CNN-LSTM model.

Model	Direction	*MAE* / g	*RMSE* / g	*TIC*	*PCC*
FNN	Vertical	0.0122	0.0156	0.6558	0.4670
	Lateral	0.0079	0.0102	0.7828	0.1271
LSTM	Vertical	0.0069	0.0088	0.2672	0.8633
	Lateral	0.0066	0.0085	0.5238	0.6036
CNN-LSTM	Vertical	0.0062	0.0079	0.2457	0.8984
	Lateral	0.0054	0.0069	0.3600	0.7387

According to the above discussions, the specific configurations of CNN-LSTM are given in Table 1.

C. MODEL ASSESSMENT

CNN-LSTM is constructed on the TensorFlow framework with the support of two GeForce GTX 1080Ti GPUs. The inference time for the testing data (30 km) is about 110 s, which is faster than the running speed of a HSR vehicle (360 s for 30 km).

Besides, another two models were also constructed to compare with CNN-LSTM. One is the FNN employed by PBTG and the other is the plain LSTM. Taking Formula (8) – (11) as assessment indices, the results are listed in Table 2.

CNN-LSTM has superior performance compared with other models. This is benefited from the combination of CNN and LSTM as well as the careful design of model configurations. To further illustrate model performance, a section of VBA waveforms and the power spectral density (PSD) of the whole testing data are given in Figure 10.

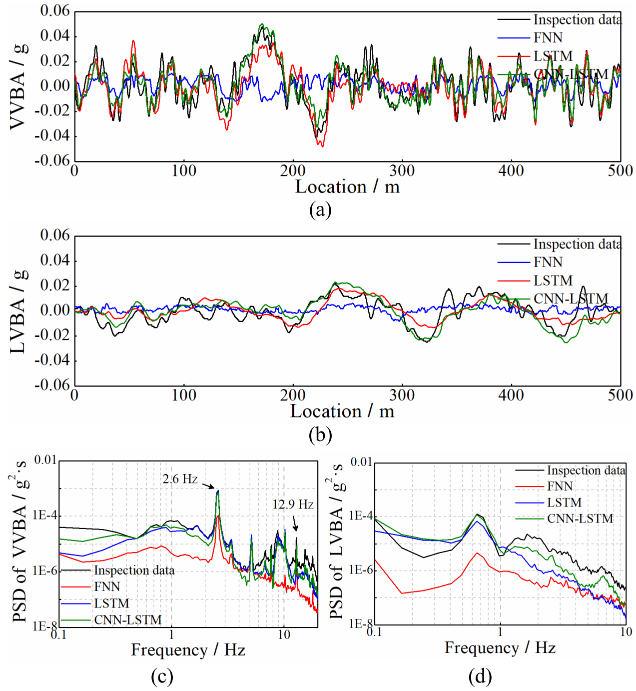

FIGURE 10. Performance comparison of different models. The line in black represents the measured VBA and the lines in color are the predicted VBA by different models. According to [3], VVBA signal is processed by 20 Hz low-pass filtering and LVBA signal is processed by 10 Hz low-pass filtering. (a) Waveform of VVBA. (b) Waveform of LVBA. (c) PSD of VVBA. (d) PSD of LVBA.

As can be seen from Figure 10, the waveforms predicted by CNN-LSTM stay closest to the inspection data. Furthermore, CNN-LSTM outperforms LSTM in predicting VVBA at frequencies below 10 Hz, and in predicting LVBA at spatial frequencies below 1 Hz. The performance gain is mainly generated from the extraordinary feature extraction capacity of CNN. Meanwhile, CNN-LSTM performs well in predicting VVBA at frequencies of 2.6 Hz and 12.9 Hz, which are activated by multi-span simply-supported girders (32 m in length) and continuously-spread track slabs (6.5 m in length), respectively. However, the VBA predicted by FNN deviates greatly from the actual value, thus FNN cannot be used for point-wise prediction.

D. VISUALIZATION OF MODEL INNER STATE

To delve into the black box characteristic of the proposed model, the activation and the hidden state of CNN and LSTM are extracted and visualized. This is helpful for gaining insight into the internal behavior of these deep learning models. A track section (120 m) of track geometry waveforms is intercepted and the corresponding activations of CNN pooling layers are extracted, as shown in Figure 11. It can be seen that the multi-channel activations are different from each other, showing that different convolutional kernels can extract different features. Besides, the activations in

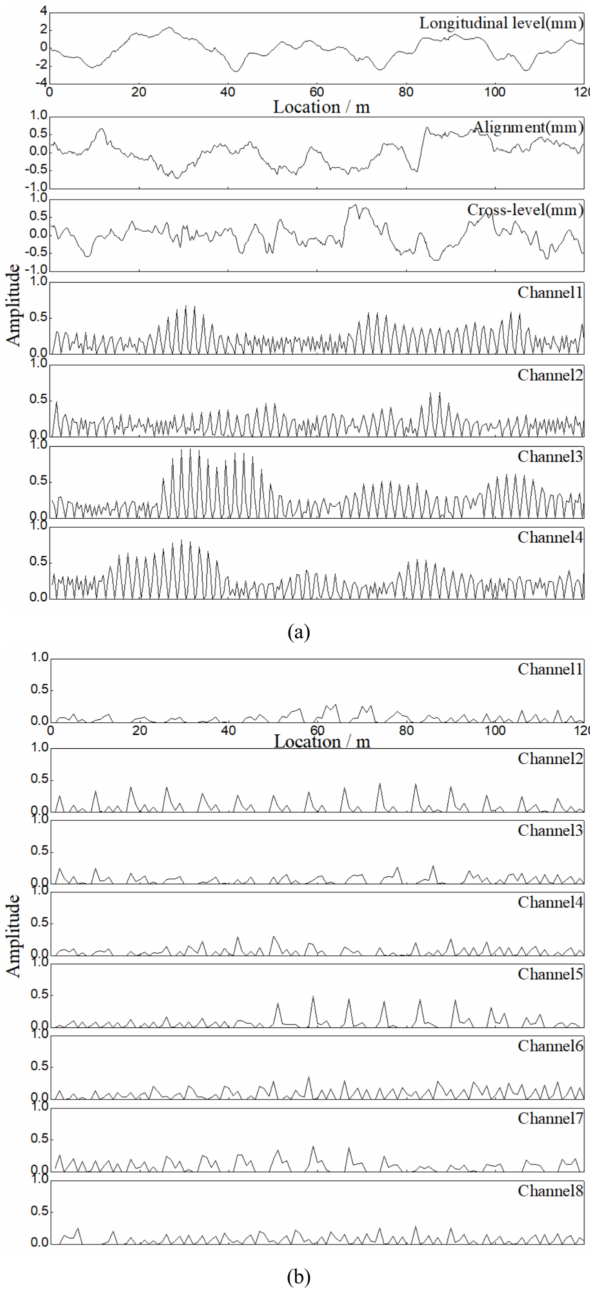

FIGURE 11. Comparison of CNN activations with track geometry waveforms. (a) Track geometry and activations of the first pooling layer. (b) Activations of the second pooling layer.

Channel 4 of the first pooling layer has similar patterns with longitudinal level, and this is a sign of shape features learning. Moreover, the activations also seem to have multi-wavelength contents, which may be learned from the multi-wavelength correlation between longitudinal level and VVBA. However, the activations of the second pooling layer are too abstract to understand, except that they are sparser than those of the first layer.

The hidden state of the top LSTM layer is plotted in Figure 12a. To facilitate comparison, the corresponding VBA waveform is also given in Figure 12b. Figure 12a is

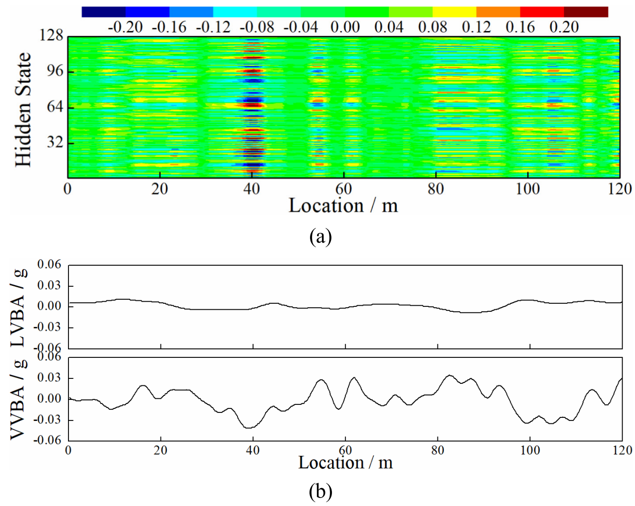

FIGURE 12. Visualization of the hidden states in the LSTM module. (a) Hidden states of the top LSTM layer. (b) VBA waveforms.

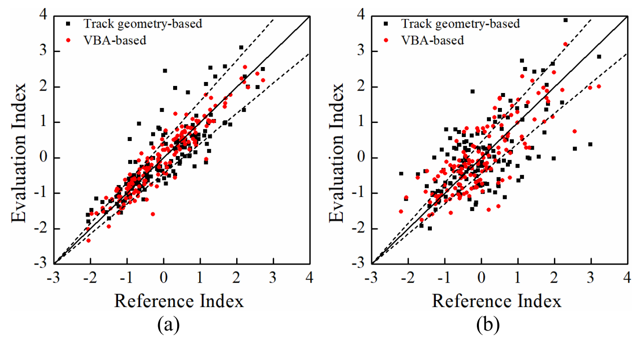

FIGURE 13. Comparison of track geometry-based and VBA-based evaluation indexes. (a) Comparison in the vertical direction. (b) Comparison in the lateral direction.

plotted from the hidden state matrix where each column represents the hidden state vector (128 elements) of a LSTM memory cell. It can be seen that the color tends to be strengthened, i.e., the hidden state vector has many large activations, where there is a peak or valley in the VVBA waveform. This indicates that LSTM can automatically learn to emphasize or suppress inner states according to the pattern of the data. Besides, it has been verified that two consecutive hidden states have higher similarity than those staying apart. This reveals that the hidden states change gradually as the information flows inside the LSTM layer, which is a sign of sequential features learning.

E. COMPARATIVE ANALYSIS

In the European Standard EN 13848-5 [1], extreme value, mean value, and standard deviation over a defined length of track geometry are taken as the evaluation index of track quality. Meanwhile, the standard defines three limit levels for evaluation indexes, including Alert Limit, Intervention Limit, and Immediate Action Limit, to guide track maintenance. Such a track geometry standard provides a convenient way to evaluate track quality. However, a track quality evaluation index should be derived from vehicle response in order for it to be "performance-based" [32]. Therefore, this paper advocates to take the VBA predicted by CNN-LSTM as an additional evaluation index.

Scatter plot can be used to compare VBA-based evaluation index and geometry-based evaluation index [8], as shown in Figure 13. In the figure, the actual VBA is taken as reference index and all the indexes are calculated by the standard deviation over a track section of 200 m (defined in EN 13848-5). Besides, the indexes are centralized for a better illustration. The solid line has slope 1 and the dashed lines are $\pm 15\%$ deviations of the solid line.

In the vertical direction, the VBA-based index is approximately proportional to the reference index. Moreover, it can be seen that the VBA-based index have a strengthened correlation with reference index, compared to track geometry-based index. In the lateral direction, however, the correlation between evaluation index and reference index is weaker than that in the vertical direction. This is mainly due to the complicated wheel-rail lateral interaction [8]. According to the above, the predicted VBA is more closely related to vehicle performance than track geometry. So that they are suggested to be taken as a performance-based track quality evaluation index on the foundation of existing track geometry standards.

IV. CONCLUSION

For the purpose of promoting track quality evaluation, a CNN-LSTM model is proposed for the point-wise prediction of HSR vehicle-body vibration by using track geometry. To achieve the premium performance, structural configurations are extensively studied. By visualizing the model inner state, it is easier to understand the working mechanism of CNN-LSTM. Analysis shows that CNN can learn shape features contained in track geometry waveform, and LSTM is capable of learning the sequential information of vehicle-body acceleration. Owing to the combined effect of these feature extraction capacities, the proposed model outperforms the fully-connected neural network adopted in PBTG and the plain LSTM in terms of accuracy. According to assessment on the inspection data of a HSR line, the root mean square error and correlation coefficient of CNN-LSTM are 0.008 g and 0.898, respectively. Besides, spectral analysis also shows that CNN-LSTM has superior performance in predicting vertical vehicle-body vibration below 10 Hz and lateral vehicle-body vibration below 1 Hz. By fine-tuning, this model can be used for predicting other vehicle responses, such as wheel-rail forces and so on. Finally, by synthesizing the predicted vehicle-body acceleration into the evaluation index of track quality, the correlation between evaluation index and vehicle performance can be enhanced. In the long run, CNN-LSTM has the potentiality in reducing the high cost of sensors installed in CITs and in promoting the health monitoring of HSR track.

However, the proposed model still needs validation by more tests for practical applications. In the future, more vehicle types and running speeds will be taken into consideration and the discrepant contributions of track geometry parameters to vehicle-body vibration will be further studied.

REFERENCES

[1] *Railway Applications-Track-Track Geometry Quality—Part 5: Geometric Quality Levels-Plain Line, Switches and Crossings*, Standard BS EN 13848-5, 2017.

[2] *Railway Applications-Testing and Simulation for the Acceptance of Running Characteristics of Railway Vehicles-Running Behavior and Stationary Tests*, Standard BS EN 14363, 2016.

[3] *Maintenance Criteria for Ballastless Track of High-Speed Railway*, Standard TG/GW 115-2012, 2012.

[4] D. Li, A. Meddah, K. Hass, and S. Kalay, "Relating track geometry to vehicle performance using neural network approach," *Proc. Inst. Mech. Eng. Part F-J. Rail Rapid Transit*, vol. 220, no. 3, pp. 273–281, 2006.

[5] L. Xu and W. Zhai, "A novel model for determining the amplitude-wavelength limits of track irregularities accompanied by a reliability assessment in railway vehicle-track dynamics," *Mech. Syst. Signal Process.*, vol. 86, pp. 260–277, Mar. 2017.

[6] W. Zhai, K. Wang, and C. Cai, "Fundamentals of vehicle-track coupled dynamics," *Veh. Syst. Dyn.*, vol. 47, no. 11, pp. 1349–1376, 2009.

[7] Y. Liu and E. Magel, "Performance-based track geometry and the track geometry interaction map," *Proc. Inst. Mech. Eng. Part F-J. Rail Rapid Transit*, vol. 223, no. 2, pp. 111–119, 2009.

[8] B. Luber, A. Haigermoser, and G. Grabner, "Track geometry evaluation method based on vehicle response prediction," *Veh. Syst. Dyn.*, vol. 48, pp. 157–173, 2010.

[9] G. M. Shafiullah, A. B. M. Ali, A. Thompson, and P. J. Wolfs, "Predicting vertical acceleration of railway wagons using regression algorithms," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 2, pp. 290–299, Jun. 2010.

[10] E. G. Berggren, M. X. D. Li, and J. Spännar, "A new approach to the analysis and presentation of vertical track geometry quality and rail roughness," *Wear*, vol. 265, no. 9, pp. 1488–1496, 2008.

[11] M. X. D. Li, E. G. Berggren, and M. B. Persson, "Assessing track geometry quality based on wavelength spectra and track-vehicle dynamic interaction," *Veh. Syst. Dyn.*, vol. 46, pp. 261–276, 2008.

[12] M. X. D. Li, E. G. Berggren, and M. Berg, "Assessment of vertical track geometry quality based on simulations of dynamic track-vehicle interaction," *Proc. Inst. Mech. Eng. F, J. Rail Rapid Transit*, vol. 223, no. 2, pp. 131–139, 2009.

[13] M. R. Minar and J. Naher, "Recent advances in deep learning: An overview," Jul. 2018, *arXiv:1807.08169*. [Online]. Available: https://arxiv.org/pdf/1807.08169.pdf

[14] K. Muhammad, J. Ahmad, I. Mehmood, S. Rho, and S. W. Baik, "Convolutional neural networks based fire detection in surveillance videos," *IEEE Access*, vol. 6, pp. 18174–18183, 2018.

[15] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, "Disease prediction by machine learning over big data from healthcare communities," *IEEE Access*, vol. 5, pp. 8869–8879, 2017.

[16] N. Neverova, C. Wolf, G. Lacey, L. Fridman, D. Chandra, B. Barbello, and G. Taylor, "Learning human identity from motion patterns," *IEEE Access*, vol. 4, pp. 1810–1820, 2016.

[17] S. Zhang, Z. Sun, J. Long, C. Li, and Y. Bai, "Dynamic condition monitoring for 3D printers by using error fusion of multiple sparse auto-encoders," *Comput. Ind.*, vol. 105, pp. 164–176, Feb. 2019.

[18] H. Shao, H. Jiang, X. Li, and S. Wu, "Intelligent fault diagnosis of rolling bearing using deep wavelet auto-encoder with extreme learning machine," *Knowl.-Based Syst.*, vol. 140, pp. 1–14, Jan. 2018.

[19] Z. He, H. Shao, X. Zhang, J. Cheng, and Y. Yang, "Improved deep transfer auto-encoder for fault diagnosis of gearbox under variable working conditions with small training samples," *IEEE Access*, vol. 7, pp. 115368–115377, 2019.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proc. ICCV*, Santiago, Chile, 2015, pp. 1026–1034.

[21] Z. Wang, W. Yan, and T. Oates, "Time series classification from scratch with deep neural networks: A strong baseline," presented at the IJCNN, Anchorage, AK, USA, 2017.

[22] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[23] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[24] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," presented at the 30th ICML, Atlanta, GA, USA, 2013.
[25] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. ICML*, Haifa, Isreal, 2010, pp. 807–814.
[26] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
[27] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," Dec. 2014, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/pdf/1412.6980.pdf
[28] M. Lin, Q. Chen, and S. Yan, "Network in network," Mar. 2013, *arXiv:1312.4400*. [Online]. Available: https://arxiv.org/pdf/1312.4400.pdf
[29] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. CVPR*, Las Vegas, NV, USA, 2016, pp. 2818–2826.
[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. CVPR*, Las Vegas, NV, USA, 2016, pp. 770–778.
[31] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," presented at the 14th ECCV, Amsterdam, The Netherlands, 2016.
[32] A. E. Fazio and J. L. Corbin, "Track quality index for high speed track," *J. Transp. Eng.*, vol. 112, no. 1, pp. 46–61, 1986.

SHUAI MA received the B.S. degree in railway engineering from Beijing Jiaotong University, Beijing, China, in 2004, where he is currently pursuing the Ph.D. degree with the School of Civil Engineering. His research interests include railway track maintenance management, track quality evaluation, track inspection data mining, and machine learning.

LIANG GAO received the B.S., M.S., and Ph.D. degrees from Southwest Jiaotong University, Chengdu, China, in 1990, 1993, and 1997, respectively.

Since 2004, he has been a Professor with the School of Civil Engineering, Beijing Jiaotong University. He is the author of 16 books and more than 180 articles, and holds more than 20 patents. His research interests include the new-type track structure, track dynamics, continuously welded rails in high-speed railways, and the structure health monitoring of the railway systems.

Prof. Gao is a member of the Committee of China Railway Corporation, the Vice Chairman of the Standardization (Track) Technical Committee of China Railway Society, and the Director of the Railway Branch of China Engineering Construction Association. He received the First Prize of The State Scientific and Technological Progress Award, in 2017, the Scientific and Technological Progress Award of The Ho Leung Ho Lee Foundation, in 2018, the National Ten-Thousand Talents Program, in 2018, the Chang Jiang Scholars Programme of Ministry of Education, in 2014, the Science and Technology Beijing 100 Leading Talents Training Project, in 2018, and the Chinese University's Top 10 Scientific and Technological Advances Award, in 2013.

XIUBO LIU received the B.S. and M.S. degrees in railway engineering from Southwest Jiaotong University, Chengdu, in 1993 and 1996, respectively, and the Ph.D. degree in highway and railway engineering from Beijing Jiaotong University, Beijing, China, in 2007.

Since 2012, he has been a Research Fellow at the Infrastructure Inspection Research Institute, China Academy of Railway Sciences Corporation Limited. His research interests include track irregularity spectrum, track irregularity assessment standards, and track maintenance management.

Dr. Liu is a member of the Standardization (Track) Technical Committee of China Railway Society and the Academic Leader of the China Railway. Since 2013, he has received five technical awards from China Railway Society.

JING LIN (M'15) received the B.S. and Ph.D. degrees from the Nanjing University of Science and Technology, Nanjing, China, in 2003 and 2008, respectively.

She is currently an Associate Professor with the Division of Operation and Maintenance Engineering, Luleå University of Technology (LTU), Luleå, Sweden. She has authored above 70 peer-reviewed journal and conference papers and one monograph in related topics. Her research interests include reliability and maintenance engineering, big data analytics, eMaintenance, and asset management.

Prof. Lin is a member of the System Engineering Society of China and the Swedish Standards Institute (SIS). She is also an associate editor of two academic journals.