Journal of Zhejiang University-SCIENCE A (Applied Physics & Engineering) 2025 26(5):424-437
www.jzus.zju.edu.cn; www.springer.com/journal/11582
E-mail: jzus_a@zju.edu.cn

Research Article

https://doi.org/10.1631/jzus.A2300643

Road pavement performance prediction using a time series long short-term memory (LSTM) model

Chuanchuan HOU[1], Huan WANG[1], Wei GUAN[2], Jun CHEN[1]✉

[1]*School of Transportation Science and Engineering, Beihang University, Beijing 100191, China*
[2]*Research Institute of Highway, Ministry of Transport, Beijing 100088, China*

Abstract: Intelligent maintenance of roads and highways requires accurate deterioration evaluation and performance prediction of asphalt pavement. To this end, we develop a time series long short-term memory (LSTM) model to predict key performance indicators (PIs) of pavement, namely the international roughness index (IRI) and rutting depth (RD). Subsequently, we propose a comprehensive performance indicator for the pavement quality index (PQI), which leverages the highway performance assessment standard method, entropy weight method, and fuzzy comprehensive evaluation method. This indicator can evaluate the overall performance condition of the pavement. The data used for the model development and analysis are extracted from tests on two full-scale accelerated test tracks, called MnRoad and RIOHTrack. Six variables are used as predictors, including temperature, precipitation, total traffic volume, asphalt surface layer thickness, pavement age, and maintenance condition. Furthermore, wavelet denoising is performed to analyze the impact of missing or abnormal data on the LSTM model accuracy. In comparison to a traditional autoregressive integrated moving average (ARIMAX) model, the proposed LSTM model performs better in terms of PI prediction and resiliency to noise. Finally, the overall prediction accuracy of our proposed performance indicator PQI is 93.8%.

Key words: Asphalt pavement performance model; International roughness index (IRI); Rutting depth (RD); Long short-term memory (LSTM) model; Pavement management system

1 Introduction

The performance of road pavement should be carefully and regularly evaluated to ensure the safety and efficiency of roads and highways. For this purpose, pavement management systems have been established to monitor and collect current road information, predict future conditions, and develop strategies for the maintenance and rehabilitation of various pavement types, such as asphalt (Kay et al., 1993; Zaghloul et al., 2006; AASHTO, 2012). In these pavement management systems, performance indicators (PIs) such as the international roughness index (IRI) and rutting depth (RD) are among the major factors used to evaluate the status of asphalt pavement. Therefore, it is vital to establish reliable prediction models for these PIs and develop condition assessment models based on the predictions.

IRI is an indicator of pavement performance developed by the World Bank in the 1980s. It is defined as the cumulative vertical displacement caused by a quarter-vehicle model driving at a speed of 80 km/h. The IRI can reflect the imperfections and usage conditions of roads and highways, helping with assessments of driving quality, driving safety, recommended driving speeds, and road maintenance needs (Sayers et al., 1986; AASHTO, 2012). Rutting, wherein longitudinal surface depressions form in a road, is another major concern for pavement maintenance. Thus, RD is generally used as the indicator of the severity of pavement rutting (Robbins and Tran, 2016). At present, many highway agencies across the world regard the initial values of these PIs (IRI and RD of the pavement immediately after construction) as quality assurance standards, and current PI values as indicators of whether the pavement needs maintenance or reconstruction. Due to the importance of such PIs for

✉ Jun CHEN, junchen@buaa.edu.cn
Chuanchuan HOU, https://orcid.org/0000-0002-0615-5439

Received Dec. 19, 2023; Revision accepted June 2, 2024;
Crosschecked Apr. 16, 2025

evaluating pavement performance, models that can predict them have garnered significant research attention.

Various types of models have been established for the prediction of pavement PIs. These models can generally be classified into deterministic and probabilistic models (AASHTO, 2012; Abaza, 2016; Justo-Silva et al., 2021). In the deterministic category, many models are based on multiple linear regression techniques. For example, different multiple linear regression models were developed to predict IRI for different pavement types across the world (Al-Suleiman and Shiyab, 2003; Albuquerque and Núñez, 2011; Owolabi et al., 2012). Although simple to construct and modify, multiple linear regression models may not accurately capture the complex relationship between pavement PIs and predicting factors. With the recent developments in artificial intelligence, researchers have begun to use machine learning methods instead of traditional regression models to predict PIs. For instance, some researchers used support vector regression (SVR) models to predict pavement performance. Li Z et al. (2021) combined the SVR with particle swarm optimization (PSO) to improve the search efficiency and parameter continuity of the SVR model, resulting in accurate prediction of rutting. Also, Zhao et al. (2022) studied the impact of traffic loading on pavement performance using SVR and found significant accuracy improvement compared to traditional methods. Furthermore, Kaloop et al. (2022) adopted a hybrid wavelet-optimally-pruned extreme learning machine (WOPELM) model to estimate the IRI of rigid pavements, resulting in better performance than several other machine learning models.

The back-propagation neural network (BPNN) algorithm has also been adopted by several researchers for pavement PI prediction (Choi et al., 2004; Gong et al., 2018; Yao et al., 2019; Abdelaziz et al., 2020). Compared to regression models, neural networks act as so-called "black boxes", with the exact relationship between the independent and predicted variables being unknown (Pérez-Acebo et al., 2020). Moreover, conventional neural networks generally consider data as isolated points in the time domain, with causality between past and future data not being leveraged. To tackle this issue, some researchers have employed time series models for pavement performance prediction. Compared to conventional neural networks, time series models consider the PIs not as isolated data points, but as time-dependent data series, where the future is dependent upon the past. For instance, a random forest algorithm was used to develop a time series model for IRI prediction after 5 and 10 years based on the long-term pavement performance (LTPP) database (Marcelino et al., 2021). Another innovative IRI prediction model using the LTPP database was developed using fuzzy-trend time series forecasting and PSO (Li et al., 2019).

Long short-term memory (LSTM) models have also been employed for time-dependent predictions. For instance, an LSTM model adopting the recurrent neural network (RNN) algorithm was developed to predict pavement deterioration based on monitoring data from the Korean National Highway Pavement Management System (Choi and Do, 2019). Additionally, a novel feature fusion LSTM-BPNN model was proposed to better capture the latent relationship between cross-sectional and time-series features through an attention mechanism (Dong et al., 2019). Recently, many advanced algorithms leveraging time series models have been proposed to improve pavement performance prediction. For example, a two-stage model was developed by Tabatabaee et al. (2013), which groups road sections with similar characteristics using a support vector classifier (SVC), and then accurately predicts the pavement performance using an RNN based on the first-stage classification results. In another study, a wavelet-autoregressive moving average (W-ARMA) model was established to reduce the influence of frequent maintenance on asphalt rutting evaluation, which filtered out noise signals caused by repair work using wavelet denoising (Fang et al., 2020).

In summary, the existing models used to predict pavement PIs can be divided into three categories: multiple linear regression models, traditional neural network models (such as BPNN), and time series models (such as RNN and LSTM). A summary of the literature on pavement PI prediction models is presented in Table 1. Since the time series models have the unique advantage of exploiting the time-dependent nature of the PIs, their predictions show the most promising results.

In this study, a time series model is developed to reliably predict pavement performance. First, data from the asphalt pavements of two major full-scale accelerated test tracks, MnRoad and RIOHTrack, are utilized to develop a generalizable model. The temperature, precipitation, total traffic volume, asphalt surface layer thickness, pavement age, and the treatment

Table 1 Summary of the pavement PI prediction models

Reference	PI	Modeling technique	Predictor	Goodness of fit
Al-Suleiman and Shiyab (2003)	IRI	MLR	AGE	N=440, R^2=0.6–0.8
Albuquerque and Núñez (2011)	IRI	MLR	CLT, TRF, PSC	N=18–27, R^2=0.87–0.94
Owolabi et al. (2012)	IRI	MLR	DIS	R^2=0.78
Choi et al. (2004)	IRI	BPNN	TRF, PSC	N=117, R^2=0.71
Gong et al. (2018)	RD	BPNN	AGE, CLT, TRF, PSC	N=88, R^2=0.867
Yao et al. (2019)	RD/IRI/SFC/CRACK	BPNN	AGE, TRF, PSC	R^2=0.85–0.90
Abdelaziz et al. (2020)	IRI	BPNN	IPI, AGE, DIS	N=2439, R^2=0.75
Marcelino et al. (2021)	IRI	Random forest	CLT, TRF, PSC	N=6–274, AVG E_{MS}=0.01–0.27 m^2/km^2
Li et al. (2019)	IRI	PSO	–	E_{MS}=0.191 m^2/km^2
Choi and Do (2019)	IRI/RD/CRACK	LSTM	CLT, TRF	N=1128–5076, R^2=0.71–0.87
Dong et al. (2019)	IRI	LSTM-BPNN	IPI, CLT, TRF, PSC	N=2243, R^2=0.867
Tabatabaee et al. (2013)	IRI	SVC-RNN	AGE, CLT, PSC, MT	R^2=0.98
Fang et al. (2020)	RD	W-ARMA	–	E_R=0.01–2.83
Kaloop et al. (2022)	IRI	WOPELM	IPI, AGE, DIS, CLT, PSC	E_{MA}=0.231 m/km

SFC, sideway force coefficient; CRACK, cracking; MLR, multiple linear regression; BPNN, back-propagation neural network; PSO, particle swarm optimization; LSTM, long short-term memory; SVC, support vector classifier; RNN, recurrent neural networks; W-ARMA, wavelet-autoregressive moving average; WOPELM, wavelet-optimally-pruned extreme learning machine; AGE, pavement age; CLT, climate; TRF, traffic; PSC, pavement structure and construction; DIS, distress; IPI, initial performance indicator; MT, maintenance; N, number of observations; R^2, coefficient of determination; AVG, average; E_{MS}, mean squared error; E_R, relative error; E_{MA}, mean absolute error

condition on the pavement are set as predictors. Then, an LSTM model is constructed to predict the variation of PIs. The robustness of the proposed model is tested by applying wavelet denoising to the original data to investigate the influence of missing and abnormal data. In addition, the performance of the LSTM model is compared with that of a traditional autoregressive integrated moving average (ARIMAX) model. Furthermore, to assess and classify the pavement performance condition, a methodology combining the highway performance assessment standard method, entropy weight method, and fuzzy comprehensive evaluation method is developed based on the predicted PIs. A comprehensive performance indicator for the pavement quality index (PQI) is calculated by applying the strictest performance evaluation method among those available.

2 Data acquisition

2.1 Acquisition of performance indicator data

To facilitate growing traffic volume and other issues, supervising agencies must improve road design, construction, and maintenance by formulating relevant technical specifications. Accordingly, it is often necessary to perform tests on pavement subjected to similar traffic loading and environmental conditions as real roads and highways, to obtain a high volume of test data in a short period. Full-scale accelerated test tracks can meet these requirements, and many countries have used them to perform tests on different scales. For our pavement performance prediction model, we use data from the mainline road research project of the Minnesota Department of Transportation, USA (MnRoad) (DOT, 2022), and the full-scale loop project of the Institute of Highway Research of the Ministry of Transport of China (RIOHTrack) (Li S et al., 2021).

The MnRoad mainline is a 3.5 mile-long (5.63 km) interstate road diverted from the parallel westbound I-94 (Interstate 94) between Albertville and Monticello, Minnesota, USA. It consists of two driving and passing lanes carrying different traffic on identical structures. In the pavement, 27 asphalt concrete test cells were constructed in 1993, and monitored until reconstruction in 2008. More details of MnRoad, such as photos of the tracks, can be found on the official website of the project (DOT, 2022). The RIOHTrack was constructed in Nov. 2015 and is

located at the highway traffic test site of the Ministry of Transportation of China in Beijing. After a year of calibration, loading tests began on the track on Nov. 28, 2016. We utilized the data of 19 main test cells from Dec. 2016 to Jan. 2020 in this study.

For the MnRoad sections, the PI data were collected once a year during different seasons for 15 years, while for the RIOHTrack sections, the data were collected once a month for 38 months. PI data from these test tracks, namely the IRI and the RD, are summarized in Fig. 1. A comparison of the descriptive statistics for the PIs of the two test tracks is presented in Table 2.

From a statistical point of view, IRI and RD evolve similarly (increasing steadily with time) except for the data of IRI in RIOHTrack, which has no obvious upward trend. However, a major micro-surfacing treatment was applied to some MnRoad sections in 1999 and 2003, resulting in improved pavement performance.

2.2 Definition of predictors

The performance of asphalt pavement can be affected by climatic conditions, repeated traffic loading, and the pavement's structure. Climate mainly influences pavement in two ways: temperature variation and water penetration. Traffic loading is one of the direct causes of road damage and plays an important role in road performance. The pavement surface layer can suffer from the vertical force, horizontal force, and impact force of heavy traffic loads, and simultaneously can be affected by precipitation, erosion, and temperature changes. However, it is important to note

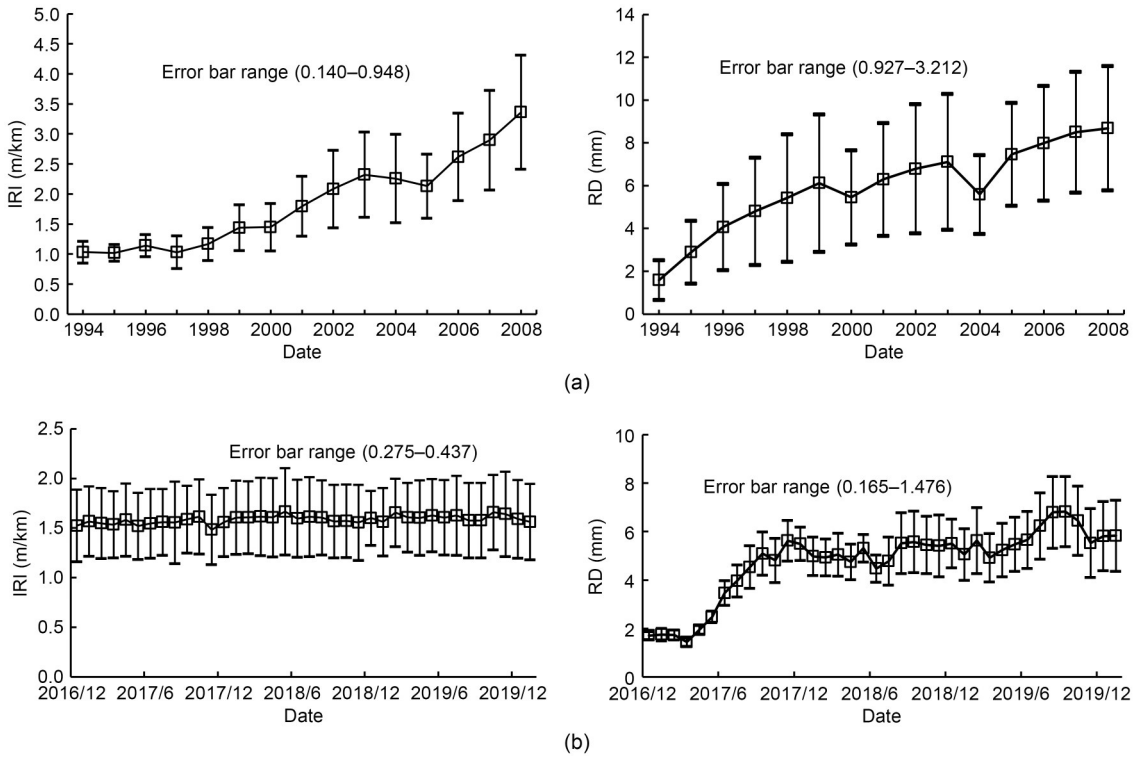

Fig. 1 PI data for pavement sections: (a) MnRoad (27 samples); (b) RIOHTrack (19 samples)

Table 2 Statistics of PIs for the two full-scale accelerated test tracks

Item	MnRoad (15 years, 27 sections)		RIOHTrack (38 months, 19 sections)	
	IRI (m/km)	RD (mm)	IRI (m/km)	RD (mm)
Average	1.849	5.797	1.586	4.769
Standard deviation	0.907	3.141	0.363	1.728
Min	0.666	0.826	0.916	1.139
Max	5.390	16.492	2.644	9.104

that treatment can significantly improve pavement performance. Therefore, in this study, we adopt the following eight predictors for our time series models (Tabatabaee et al., 2013; Choi and Do, 2019):

1. TEMP_AVE (℃)—the average temperature from the moment of the previous measurement until the present moment;

2. TEMP_MAX (℃)—the maximum temperature from the moment of the previous measurement until the present moment;

3. TEMP_MIN (℃)—the minimum temperature from the moment of the previous measurement until the present moment;

4. PRECIP (mm)—the total precipitation from the moment of the previous measurement until the present moment;

5. TRAFFIC (ESAL)—the total traffic load from the moment of the previous measurement until the present moment, measured using the equivalent single axle load (ESAL) (Alavi and Senn, 1999);

6. AC_thickness (mm)—the asphalt surface layer thickness;

7. AGE (year/month)—inspection year (for MnRoad)/month (for RIOHTrack) minus construction year/month;

8. Treatment—a dummy variable (1 indicates treatment applied in this year, and 0 indicates otherwise).

Note that the time period for the above definitions is from the beginning to the end of a specified year for MnRoad, and from the beginning to the end of a specified month for RIOHTrack.

For MnRoad sections, some maintenance was conducted between 1993 and 2008, so the dummy variable may not be 0 during that time. But for RIOHTrack sections, the TRAFFIC, AC_thickness, and dummy variable from the moment of the previous measurement until the present moment remained unchanged during the data collection period; therefore we do not consider these variables as predictors. Among the predictors, we show the curves of TEMP, PRECIP, and TRAFFIC over time for the two full-scale accelerated test tracks in Fig. 2, where TRAFFIC consists of two different lanes (a driving lane and a passing lane).

3 Wavelet denoising of original data

Intuitively, the performance of pavement surfaces worsens over time. However, measurement errors also result in anomalous changes in PIs. Therefore, we apply wavelet denoising to the original data to reduce the noise caused by measurement error, allowing a more accurate analysis of the results.

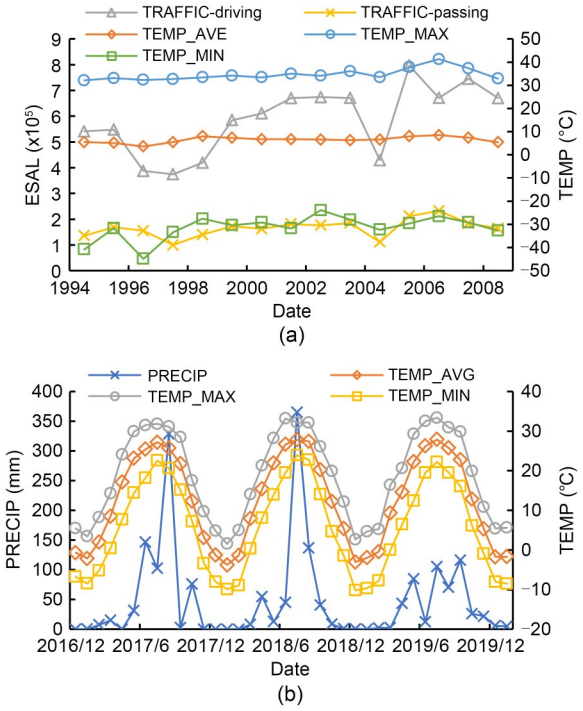

Fig. 2 Representative predictors for pavement sections: (a) MnRoad; (b) RIOHTrack

The wavelet transform is a localized analysis of temporal (or spatial) frequency that can highlight certain features of raw data. The workflow of wavelet denoising is shown in Fig. 3. In this process, the chosen wavelet algorithm, number of decomposition layers, threshold calculation algorithm, and threshold processing method all can affect the final denoising result. These factors are each discussed as follows:

1. The selection of the wavelet algorithm should consider the support length, vanishing moment, symmetry, regularity, and similarity. Since different wavelet algorithms have different processing characteristics, no single wavelet algorithm can achieve the best denoising effect for all types of signals. However, the Daubechies (DB) and the Symlet (SYM) algorithms are two families of wavelet bases that generally work well for denoising (Stanke et al., 2020).

2. In wavelet decomposition, the choice of the number of decomposition layers is also impactful. As the number of decomposition layers increases, the characteristic difference between noise and real signal

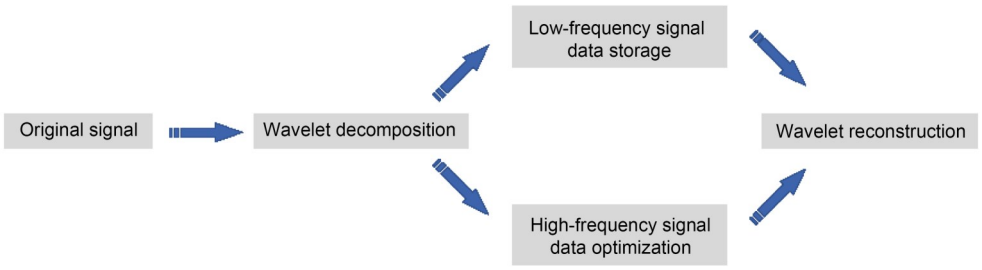

Fig. 3 Process of wavelet denoising

is more obvious, which enables better noise filtering. On the other hand, the distortion of the reconstructed signal becomes more pronounced, which will affect the final denoised result. Therefore, an appropriate number of decomposition layers should be selected to balance these factors.

3. In the process of denoising, one must take care to not remove valid information. Thus, a threshold must be appropriately selected to minimize impact on the true signal. Commonly used algorithms for threshold calculation include Rigrsure, Heursure, Sqtwolog, and Minimaxi (Valencia et al., 2016).

4. Following wavelet decomposition, an appropriate threshold processing method needs to be selected to denoise the data. Soft and hard threshold functions are often used for this purpose. The hard threshold function results in a denoised signal that is closer to the actual signal, but the signal will have added oscillations and spikes. In contrast, the soft threshold function results in a denoised signal with better overall continuity.

4 Methodology

4.1 ARIMAX time series model

The autoregressive integrated moving average (ARIMA) model is a famous time series prediction method incorporating the autoregression (AR) of lagged time series values, the transformation of the data to a stationary form ("integrated", abbreviated as "I"), and the moving average (MA) of lagged forecast errors (Box and Jenkins, 1970). An ARIMA model is specified by three parameters: p, q, and d, where p is the autoregressive term, q is the moving average term, and d is the number of subtractions applied to the lagged indicators to achieve stationary data. In this study, we adopt a revised ARIMA model that employs multiple input time series data, which we refer to as an ARIMAX model. The parameters p and q in ARIMAX are determined by the Bayesian information criterion (B) as shown in Eq. (1) (Schwarz, 1978):

$$B = k\ln n - 2\ln L, \tag{1}$$

where k is the number of parameters, n is the number of samples, and L is the maximum likelihood function value. When B is minimized, we choose the corresponding p and q values. Fig. 4 illustrates the modeling process.

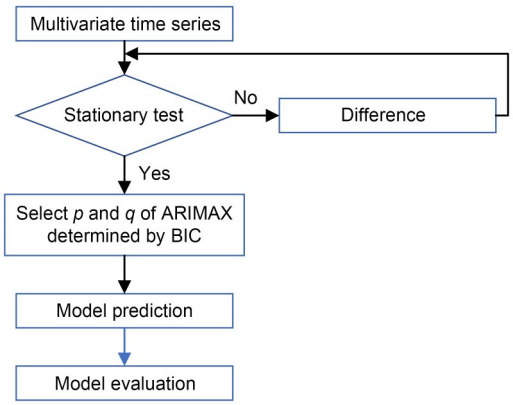

Fig. 4 Modeling process for ARIMAX. BIC is the Bayesian information criterion

4.2 LSTM

Deep learning aims at developing neural networks to process or analyze various types of data in a similar manner to the human brain, such as text, audio, images, and videos (Hinton and Salakhutdinov, 2006). Among deep learning models, RNNs are frequently used to process time series data. Compared to BPNNs and convolutional neural networks (CNNs), RNNs work by assuming that human cognition is based on experience and memory. Beyond just using

present input data, these networks utilize a memory function to harness previous information.

Although RNNs are often effective in processing time series data, they can still exhibit problems of vanishing and exploding gradients, which are caused by long-term time dependence. Thus, the LSTM model was developed to solve these problems, and is currently widely used in many fields (Dong et al., 2019). A schematic of the LSTM is shown in Fig. 5. The cell state (in the orange dashed frame) is the core of the LSTM model, as it carries the information of all previous states, and facilitates corresponding operations that decide whether to discard or add information. In addition, there are gates applied to control the input and output of information; these are the forget gate (red dashed frame), input gate (green dashed frame), and output gate (blue dashed frame). The forget gate determines what information is discarded, the input gate determines what new information is added to the cell state, and the output gate determines what information to output from the cell state.

We propose an LSTM time series model for pavement PI prediction, developed using the Keras library in Python. The time series data were preprocessed according to the approach developed by Bontempi et al. (2012). In this study, we treat the time series prediction as a supervised learning task. First, the maximum and minimum values in the training dataset are used to normalize all data to values between 0 and 1. Afterward, the time series data are transformed into supervised learning data with features and labels. The training dataset is constructed as follows: for the MnRoad sections, the input length is 1 and the output length is 1, indicating that the pavement performance during year t+1 is predicted based on the pavement performance and predictors during year t. For the RIOHTrack sections, the input length is 5 and the output length is 2, indicating that the pavement performance during months t+1 and t+2 is predicted based on the performance and predictors of the past five months (months t−4 to t).

5 Results and discussion

5.1 Model evaluation

Three regression indicators are used to assess the performance of the time series models. These indicators are the mean squared error (E_{MS}), mean absolute error (E_{MA}), and coefficient of determination (R^2), whose formulas are given in Eqs. (2)–(4):

$$E_{MS}=\frac{1}{N}\sum_{i=1}^{N}(y_i-f_i)^2, \tag{2}$$

$$E_{MA}=\frac{1}{N}\sum_{i=1}^{N}|y_i-f_i|, \tag{3}$$

$$R^2=1-\frac{\sum_{i=1}^{N}(y_i-f_i)^2}{\sum_{i=1}^{N}(y_i-\bar{y})^2}, \tag{4}$$

where N is the number of samples, y_i and f_i are the measured and predicted data, respectively, and $\bar{y}$ is the average of the measured data.

In this study, we used measured PIs of MnRoad sections from 1994 to 2007, and PIs of RIOHTrack sections from Dec. 2016 to Nov. 2019 as training data for the time series models. MnRoad section data from 2008, and RIOHTrack section data from Dec. 2019 and Jan. 2020 are used as the testing data.

The performance of the ARIMAX time series model is presented in Fig. 6, comparing the predicted

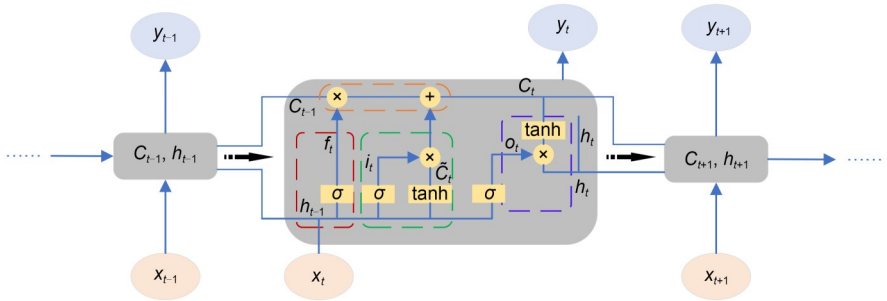

Fig. 5 Schematic of the LSTM model. C_t is the cell state at time t, h_t is the hidden state of the layer at time t, x_t is the input at time t, y_t is the output at time t, f_t is the forget gate, i_t is the input gate, $\tilde{C}_t$ is the cell gate, o_t is the output gate, σ is the sigmoid function. References to color refer to the online version of this figure

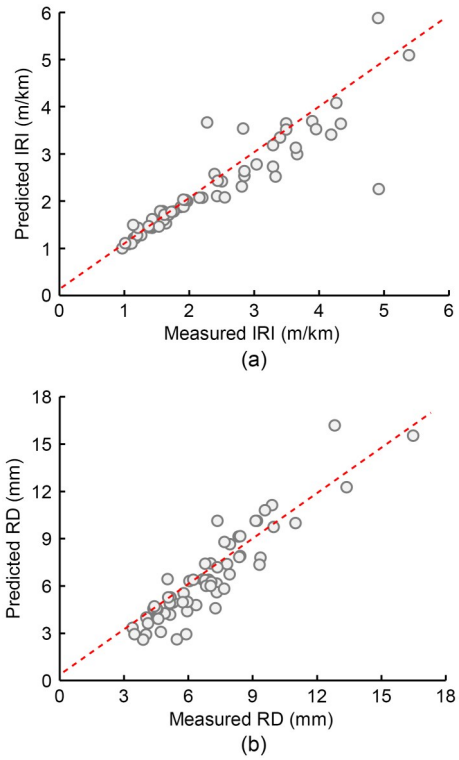

Fig. 6 Performance of the ARIMAX model for raw PI prediction: (a) IRI; (b) RD

and measured PI data. Because the predicted IRI and RD on the two roads have the same physical meanings and are within similar ranges, we combine the respective IRI and RD results across the roads. Table 3 summarizes the statistical indicators of the model performance for the MnRoad sections, the RIOHTrack sections, and the overall datasets. We can see that the ARIMAX model has a low prediction accuracy for PIs, except for the IRI of RIOHTrack, where R^2 is 0.924, E_{MS} is 0.011 m^2/km^2, and E_{MA} is 0.074 m/km. This is because the age of RIOHTrack is young, and its IRI results have no obvious upward trend (as shown in Fig. 1b), so it is relatively easy to predict.

Table 3 Performance evaluation of the ARIMAX model

Dataset	PI	Performance evaluation indicator		
		R^2	E_{MS}	E_{MA}
MnRoad	IRI	0.375	0.541 m^2/km^2	0.511 m/km
	RD	0.600	3.438 mm^2	1.258 mm
RIOHTrack	IRI	0.924	0.011 m^2/km^2	0.074 m/km
	RD	0.395	1.200 mm^2	0.840 mm
Overall	IRI	0.810	0.231 m^2/km^2	0.255 m/km
	RD	0.675	2.128 mm^2	1.014 mm

We employ a stacked LSTM model, which combines two LSTM layers for encoding sequential information from the input, and a dense layer for outputting a vector from the second LSTM layer. The size of the output vector is equivalent to the number of steps to be predicted. The E_{MS} is used as the loss function along with Adam optimization during training. We also continuously reduce the learning rate during the training process, to take smaller steps that approach a minimum in the loss function. In training, approximately 50 epochs are required to reach convergence.

Fig. 7 presents the correlation of predicted PIs with measured data using the LSTM model. Table 4 summarizes LSTM model performance evaluations across different datasets. By comparing Table 3 with Table 4, it is clear that the performance of the LSTM model is significantly better than the traditional ARIMAX model; the coefficient of determination of the LSTM model is on average 21.0% higher than that of the ARIMAX model, and the E_{MS} and E_{MA} of LSTM are 0.182 m^2/km^2 and 0.114 m/km lower than those of the ARIMAX model for IRI, and 1.772 mm^2 and

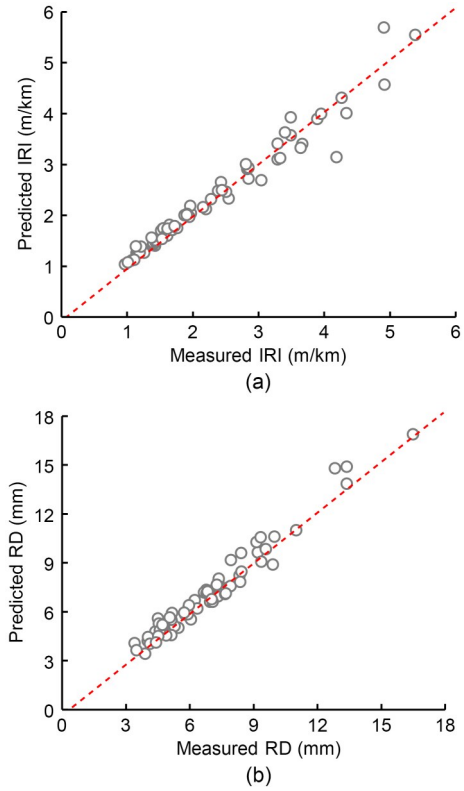

Fig. 7 Performance of the LSTM model for raw PI prediction: (a) IRI; (b) RD

Table 4 Performance evaluation of the LSTM model

Dataset	PI	Performance evaluation indicator		
		R^2	E_{MS}	E_{MA}
MnRoad	IRI	0.878	0.105 m²/km²	0.235 m/km
	RD	0.927	0.624 mm²	0.637 mm
RIOHTrack	IRI	0.932	0.010 m²/km²	0.075 m/km
	RD	0.916	0.165 mm²	0.363 mm
Overall	IRI	0.959	0.049 m²/km²	0.141 m/km
	RD	0.946	0.356 mm²	0.477 mm

0.537 mm lower for RD, respectively. This is because the ARIMAX model assumes that variables tend to vary linearly with time, and thus has weaker prediction capability on datasets with more complex time-varying behavior. In contrast, the LSTM model is more effective at learning the correlations, collinearities, and nonlinear relationships between variables.

As shown in Fig. 1, variations in curves of the original PI data from the MnRoad and RIOHTrack sections are discrete, indicating the presence of noise. Thus, we apply the aforementioned wavelet denoising method to first process the original data, enabling analysis of the influence of missing or abnormal data on the model accuracy.

The Rigrsure algorithm and soft threshold function were adopted in the wavelet denoising procedure (Dong et al., 2019). Therefore, the main factors affecting the final denoising effect are the selection of the wavelet algorithm and the number of decomposition layers. The optimal combination of wavelet factors was obtained by conducting different trials using the Daubechies order-4 (DB4), Symlet order-4 (SYM4), and Haar wavelet algorithms, along with either 3, 4, or 5 decomposition layers. The root mean square error (E_{RMS}) and smoothness (r) are calculated according to Eqs. (5) and (6) to evaluate the denoising performance:

$$E_{RMS}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(s_i-f_i)^2}, \tag{5}$$

$$r=\frac{\sum_{i=1}^{N-1}(f_{i+1}-f_i)}{\sum_{i=1}^{N-1}(s_{i+1}-s_i)}, \tag{6}$$

where s_i and f_i represent the original and the denoised data, respectively. The DB4 algorithm with three wavelet layers is selected, since it achieved the best values of root mean square error and smoothness as shown in Table 5. Fig. 8 shows the raw data of PIs from MnRoad pavement Section 1, and the data after wavelet denoising. Clearly, the data become much smoother following denoising.

Table 5 Wavelet denoising performance on the original data of RD from MnRoad Section 1

Wavelet algorithm	Number of layers	E_{RMS} (mm)	r
Haar	3	0.525	0.686
	4	0.886	0.426
	5	0.886	0.426
DB4	3	0.378	0.742
	4	0.413	0.680
	5	0.435	0.654
SYM4	3	0.397	0.672
	4	0.428	0.605
	5	0.445	0.583

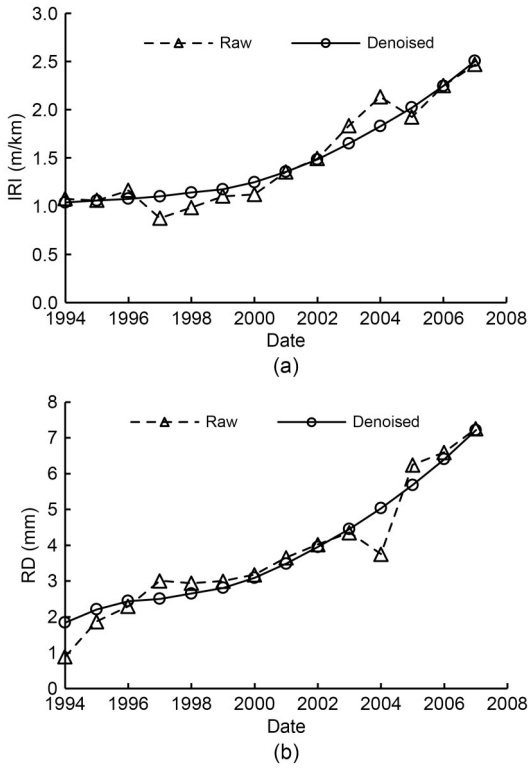

Fig. 8 Comparison of measured and denoised PI values from MnRoad Section 1: (a) IRI; (b) RD

After performing wavelet denoising on the original data, the ARIMAX and LSTM models are reapplied to predict pavement performance. The prediction

results are summarized in Table 6. The prediction accuracy of the ARIMAX model using denoised PI data is better than that using raw data, on average increasing the determination coefficient by 9.5% and decreasing the E_{MS} and E_{MA} by 0.155 m^2/km^2 and 0.087 m/km for IRI, and 0.403 mm^2 and 0.238 mm for RD, respectively. For the LSTM model, the wavelet denoising process resulted in increased RD prediction accuracy, with the determination coefficient increasing by 1.4%, E_{MS} having a slight increase of 0.012 m^2/km^2 and E_{MA} decreasing by 0.003 m/km for IRI, and E_{MS} and E_{MA} decreasing by 0.097 mm^2 and 0.098 mm for RD, respectively. Thus, wavelet denoising can significantly improve the accuracy of the ARIMAX model, and the LSTM model had a stronger tolerance to noise in the original data compared to the ARIMAX model.

It is worth noting that for the LSTM model, the wavelet denoising could not improve the accuracy of the IRI prediction. This may be because although the wavelet denoising method can reduce the noise in the original data, it may also distort the measured data. Compared to noise, distortions in the data can have a greater impact on the prediction accuracy of the LSTM model. A comparison between the measured and predicted data for the combined dataset, using the two different models, is shown in Fig. 9. Since the proposed LSTM prediction model has a strong tolerance

Table 6 Performance evaluation for denoised data

Dataset	Model	IRI			RD		
		R^2	E_{MS} (m^2/km^2)	E_{MA} (m/km)	R^2	E_{MS} (mm^2)	E_{MA} (mm)
MnRoad	LSTM	0.841	0.138	0.249	0.945	0.471	0.527
	ARIMAX	0.801	0.173	0.310	0.630	3.162	1.032
RIOHTrack	LSTM	0.952	0.007	0.061	0.946	0.108	0.274
	ARIMAX	0.951	0.007	0.068	0.644	0.705	0.594
Overall	LSTM	0.950	0.061	0.139	0.960	0.259	0.379
	ARIMAX	0.938	0.076	0.168	0.736	1.725	0.776

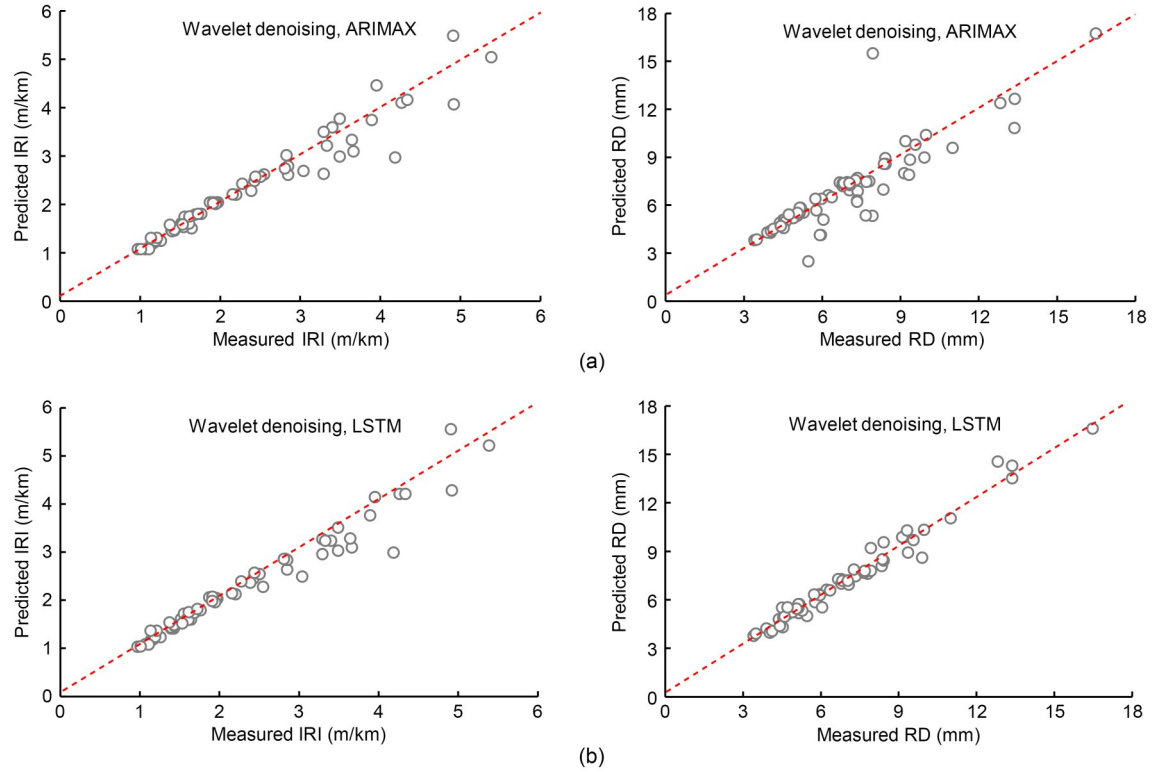

Fig. 9 Model performance for denoised PI prediction: (a) ARIMAX; (b) LSTM

to noise, and because we aim to avoid distortion, the original data should be used for LSTM prediction.

5.2 Comprehensive performance indicator prediction

The proposed LSTM model accurately predicted the pavement PIs. However, it should be noted that different PIs have different measurement units and orders of magnitude, which makes it challenging to make assessments with the predicted results. In order to more easily compare the performance of different pavements, a comprehensive performance indicator, PQI, is introduced, which depends on the value of each pavement PI. We calculate PQI based on the predicted PIs as shown in Eq. (7):

$$\begin{aligned} &I_{\mathrm{PQ}}=\omega_{\mathrm{RQ}}\times I_{\mathrm{RQ}}+\omega_{\mathrm{RD}}\times I_{\mathrm{RD}},\\ &I_{\mathrm{RQ}}=\frac{100}{1+0.026\mathrm{e}^{0.65I_{\mathrm{IR}}}},\\ &I_{\mathrm{RD}}=\begin{cases}100-d_{\mathrm{R}}, & d_{\mathrm{R}}\leqslant 10,\\ 90-3(d_{\mathrm{R}}-10), & 10<d_{\mathrm{R}}\leqslant 40,\\ 0, & d_{\mathrm{R}}>40,\end{cases}\end{aligned} \tag{7}$$

where I_{PQ} is the pavement quality index, I_{RQ} is the pavement riding quality index, I_{RD} is the pavement rutting depth index, and ω_{RQ} and ω_{RD} are the weights of I_{RQ} and I_{RD}, respectively. I_{IR} is the international roughness index, and d_{R} is the rutting depth.

To make the comparison more intuitive, the calculated values are converted to discrete numbers representing different performance states. Since the RIOHTrack pavement sections in the dataset were built in 2015 and the pavement performance had not severely degraded, we only use three performance states in this study: state 1 represents the best performance condition of the pavement, and state 3 is the worst condition. The classification standards are shown in Table 7.

Table 7 Classification standards of evaluation indicators

State	PQI
1	$I_{\mathrm{PQ}}\geqslant 90$
2	$80\leqslant I_{\mathrm{PQ}}<90$
3	$I_{\mathrm{PQ}}<80$

We utilize several different versions of PQI for our analysis:

1. PQI-S: PQI is calculated according to the assessment standard method.

2. PQI-E: PQI is calculated according to the entropy weight method.

3. PQI-SF: PQI is calculated by combining the fuzzy comprehensive evaluation method with the assessment standard method.

4. PQI-EF: PQI is calculated by combining the fuzzy comprehensive evaluation method with the entropy weight method.

5. PQI-F: PQI is determined by applying the strictest pavement performance evaluation method above.

The measured PIs and calculated PQI classifications from the training set are used to train a BPNN, and then the predicted PIs for the test set are fed into the trained network to obtain predicted PQI results. Fig. 10 shows the confusion matrix of the BPNN prediction model, where the diagonal shows the numbers of correct predictions for the three classes. The confusion matrix shows an overall prediction accuracy of 93.8%, indicating effective comprehensive performance prediction. To identify potential causes of the incorrect predictions, the measured values and predicted values across different indicators are listed in Table 8. One can observe that incorrectly predicted PQIs are around the critical value of state transition for pavement sections. Another possible reason for prediction errors is the abnormal increase in IRI for MnRoad sections in 2008, causing the predicted values to be lower than the measured values.

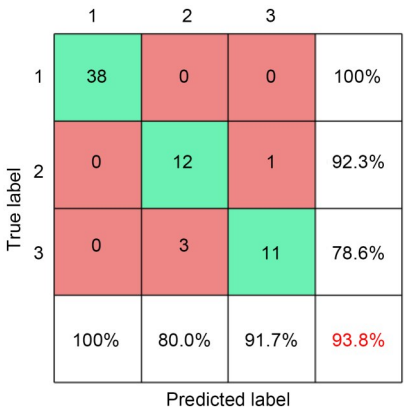

Fig. 10 Confusion matrix of the BPNN prediction model for comprehensive performance

6 Conclusions

A method based on a time series LSTM model was developed to predict the PIs of asphalt pavements

Table 8 Performance evaluation for denoised data

Data	No.	IRI (m/km)	RD (mm)	PQI-S		PQI-E		PQI-SF state	PQI-EF state	PQI-F state
				Value	State	Value	State			
Measured	1	3.67	8.35	82.50	2	86.95	2	3	1	3
	2	3.65	5.16	83.71	2	89.11	2	3	1	3
	3	4.19	9.20	77.95	3	84.19	2	3	1	3
	4	3.41	7.36	84.69	2	88.55	2	2	1	2
Predicted	1	3.40	8.08	84.53	2	88.12	2	2	1	2
	2	3.32	5.15	85.97	2	90.28	1	2	1	2
	3	3.14	9.86	85.56	2	87.79	2	2	2	2
	4	3.62	7.64	83.06	2	87.58	2	3	1	3

(specifically IRI and RD), and a comprehensive performance indicator PQI was proposed to evaluate the condition of the pavement. The average temperature, precipitation, total traffic volume, asphalt surface layer thickness, pavement age, and treatment condition were taken as predictors, and wavelet denoising was performed on the original data to investigate the influence of missing or abnormal data on the model. The modeling and analysis were based on data from two major full-scale accelerated test tracks, MnRoad and RIOHTrack, and the results were compared to that of a traditional ARIMAX model.

A comparative evaluation of the different models showed that the LSTM model outperforms the ARIMAX model in PI prediction. The coefficient of determination of the LSTM model is on average 21.0% higher than that of the ARIMAX model, and the E_{MS} and E_{MA} of LSTM are 0.182 m^2/km^2 and 0.114 m/km lower than those of the ARIMAX model for IRI, and 1.772 mm^2 and 0.537 mm lower for RD, respectively.

We also found that wavelet denoising can improve the prediction accuracy of the ARIMAX model, and that the LSTM model has a stronger tolerance to noise in the original data compared with the ARIMAX model. However, distorted data caused by the denoising process have a greater impact on the prediction accuracy of the LSTM model. Therefore, the LSTM model is particularly suitable for application scenarios with missing/noisy data or frequent maintenance operations.

Additionally, a novel comprehensive pavement performance indicator PQI was proposed in this study, which combines the highway performance assessment standard method, entropy weight method, and fuzzy comprehensive evaluation method. The overall prediction accuracy of the PQI can reach as high as 93.8%. This demonstrates the effectiveness of the method in predicting pavement performance. The findings of this study may serve as a reference for more sophisticated pavement management and maintenance strategies.

Acknowledgments

This work is supported by the National Key Research and Development Program of China (No. 2021YFB2600300).

Author contributions

Jun CHEN and Chuanchuan HOU designed the research. Huan WANG and Wei GUAN processed the corresponding data. Chuanchuan HOU and Huan WANG wrote the first draft of the manuscript. Huan WANG helped to organize the manuscript. Jun CHEN revised and edited the final version.

Conflict of interest

Chuanchuan HOU, Huan WANG, Wei GUAN, and Jun CHEN declare that they have no conflict of interest.

References

AASHTO (American Association of State Highway and Transportation Officials), 2012. Pavement Management Guide. 2nd Edition. AASHTO, Washington, USA.

Abaza KA, 2016. Back-calculation of transition probabilities for Markovian-based pavement performance prediction models. *International Journal of Pavement Engineering*, 17(3):253-264. https://doi.org/10.1080/10298436.2014.993185

Abdelaziz N, Abd El-Hakim RT, El-Badawy SM, et al., 2020. International roughness index prediction model for flexible pavements. *International Journal of Pavement Engineering*, 21(1):88-99. https://doi.org/10.1080/10298436.2018.1441414

Alavi SH, Senn KA, 1999. Development of New Pavement Design Equivalent Single Axle Load (ESAL). Report No. FHWA-AZ-99-455, Arizona Department of Transportation, Phoenix, USA.

Albuquerque FS, Núñez WP, 2011. Development of roughness prediction models for low-volume road networks in

northeast Brazil. *Transportation Research Record: Journal of the Transportation Research Board*, 2205(1):198-205. https://doi.org/10.3141/2205-25

Al-Suleiman TI, Shiyab AMS, 2003. Prediction of pavement remaining service life using roughness data—case study in Dubai. *International Journal of Pavement Engineering*, 4(2):121-129. https://doi.org/10.1080/10298430310001634834

Bontempi G, Taieb SB, le Borgne YA, 2012. Machine Learning Strategies for Time Series Forecasting. *In*: Aufaure MA, Zimányi E (Eds.), Business Intelligence. Springer, Berlin, Germany. https://doi.org/10.1007/978-3-642-36318-4_3

Box GEP, Jenkins GM, 1970. Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco, USA.

Choi JH, Adams TM, Bahia HU, 2004. Pavement roughness modeling using back-propagation neural networks. *Computer-Aided Civil and Infrastructure Engineering*, 19(4):295-303. https://doi.org/10.1111/j.1467-8667.2004.00356.x

Choi S, Do M, 2019. Development of the road pavement deterioration model based on the deep learning method. *Electronics*, 9(1):3. https://doi.org/10.3390/electronics9010003

Dong YS, Shao YX, Li XT, et al., 2019. Forecasting pavement performance with a feature fusion LSTM-BPNN model. Proceedings of the 28th ACM International Conference on Information and Knowledge Management, p.1953-1962. https://doi.org/10.1145/3357384.3357867

DOT (U.S. Department of Transportation), 2022. MnRoad Mainline–Interstate 94. Federal Highway Administration. https://infopave.fhwa.dot.gov/MnRoad/Map_Main#

Fang MJ, Han CJ, Xiao Y, et al., 2020. Prediction modelling of rutting depth index for asphalt pavement using denoising method. *International Journal of Pavement Engineering*, 21(7):895-907. https://doi.org/10.1080/10298436.2018.1512712

Gong HR, Sun YR, Mei ZJ, et al., 2018. Improving accuracy of rutting prediction for mechanistic-empirical pavement design guide with deep neural networks. *Construction and Building Materials*, 190:710-718. https://doi.org/10.1016/j.conbuildmat.2018.09.087

Hinton GE, Salakhutdinov RR, 2006. Reducing the dimensionality of data with neural networks. *Science*, 313(5786): 504-507. https://doi.org/10.1126/science.1127647

Justo-Silva R, Ferreira A, Flintsch G, 2021. Review on machine learning techniques for developing pavement performance prediction models. *Sustainability*, 13(9):5248. https://doi.org/10.3390/su13095248

Kaloop MR, El-Badawy SM, Ahn J, et al., 2022. A hybrid wavelet-optimally-pruned extreme learning machine model for the estimation of international roughness index of rigid pavements. *International Journal of Pavement Engineering*, 23(3):862-876. https://doi.org/10.1080/10298436.2020.1776281

Kay RK, Mahoney JP, Jackson NC, 1993. The WSDOT Pavement Management System—a 1993 Update. Report No. WA-RD 274.1, Washington State Department of Transportation, Olympia, USA.

Li S, Fan MM, Xu LK, et al., 2021. Rutting performance of semi-rigid base pavement in RIOHTrack and laboratory evaluation. *Frontiers in Materials*, 7:590604. https://doi.org/10.3389/fmats.2020.590604

Li W, Ju HY, Xiao LY, et al., 2019. International roughness index prediction based on multigranularity fuzzy time series and particle swarm optimization. *Expert Systems with Applications: X*, 2:100006. https://doi.org/10.1016/j.eswax.2019.100006

Li Z, Zhang JP, Liu T, et al., 2021. Using PSO-SVR algorithm to predict asphalt pavement performance. *Journal of Performance of Constructed Facilities*, 35(6):04021094. https://doi.org/10.1061/(ASCE)CF.1943-5509.0001666

Marcelino P, de Lurdes Antunes M, Fortunato E, et al., 2021. Machine learning approach for pavement performance prediction. *International Journal of Pavement Engineering*, 22(3):341-354. https://doi.org/10.1080/10298436.2019.1609673

Owolabi AO, Sadiq OM, Abiola OS, 2012. Development of performance models for a typical flexible road pavement in Nigeria. *International Journal for Traffic and Transport Engineering*, 2(3):178-184. https://doi.org/10.7708/ijtte.2012.2(3).02

Pérez-Acebo H, Linares-Unamunzaga A, Roji E, et al., 2020. IRI performance models for flexible pavements in two-lane roads until first maintenance and/or rehabilitation work. *Coatings*, 10(2):97. https://doi.org/10.3390/coatings10020097

Robbins MM, Tran NH, 2016. A Synthesis Report: Value of Pavement Smoothness and Ride Quality to Roadway Users and the Impact of Pavement Roughness on Vehicle Operating Costs. NCAT Report No. 16-03, National Center for Asphalt Technology at Auburn University, Auburn, USA.

Sayers MW, Gillespie TD, Queiroz CA, 1986. International Experiment to Establish Correlations and Standard Calibration Methods for Road Roughness Measurement. Technical Report No. 45, World Bank, Washington, USA.

Schwarz G, 1978. Estimating the dimension of a model. *Annals of Statistics*, 6(2):461-464. https://doi.org/10.1214/aos/1176344136

Stanke L, Kubicek J, Vilimek D, et al., 2020. Towards to optimal wavelet denoising scheme–a novel spatial and volumetric mapping of wavelet-based biomedical data smoothing. *Sensors*, 20(18):5301. https://doi.org/10.3390/s20185301

Tabatabaee N, Ziyadi M, Shafahi Y, 2013. Two-stage support vector classifier and recurrent neural network predictor for pavement performance modeling. *Journal of Infrastructure Systems*, 19(3):266-274. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000132

Valencia D, Orejuela D, Salazar J, et al., 2016. Comparison analysis between Rigrsure, Sqtwolog, Heursure and Minimaxi techniques using hard and soft thresholding methods. Proceedings of the 21st Symposium on Signal Processing,

Images and Artificial Vision. https://doi.org/10.1109/STSIVA.2016.7743309

Yao LY, Dong Q, Jiang JW, et al., 2019. Establishment of prediction models of asphalt pavement performance based on a novel data calibration method and neural network. *Transportation Research Record: Journal of the Transportation Research Board*, 2673(1):66-82. https://doi.org/10.1177/0361198118822501

Zaghloul S, Helali K, Bekheet W, 2006. Development and Implementation of Arizona Department of Transportation (ADOT) Pavement Management System (PMS). Report No. FHWA-AZ-06-494, Arizona Department of Transportation, Phoenix, USA.

Zhao JN, Wang H, Lu P, 2022. Impact analysis of traffic loading on pavement performance using support vector regression model. *International Journal of Pavement Engineering*, 23(11):3716-3728. https://doi.org/10.1080/10298436.2021.1915493