A Transfer Learning-Based LSTM for Traffic Flow Prediction with Missing Data

Article *in* Journal of Transportation Engineering Part A Systems · July 2023

DOI: 10.1061/JTEPBS.TEENG-7638

Zhao Zhang
McMaster University

A Transfer Learning–Based LSTM for Traffic Flow Prediction with Missing Data

Zhao Zhang, Ph.D.[1]; Hao Yang, Ph.D.[2]; and Xianfeng Yang, Ph.D.[3]

Abstract: Traffic flow prediction plays an important role in intelligent transportation systems (ITS) on freeways. However, incomplete traffic information tends to be collected by traffic detectors, which is a major constraint for existing methods to get precise traffic predictions. To overcome this limitation, this study aims to propose and evaluate a new advanced model, named transfer learning–based long short-term memory (LSTM) model for traffic flow forecasting with incomplete traffic information, that adopts traffic information from similar locations for the target location to increase the data quality. More specifically, dynamic time warping (DTW) is used to evaluate the similarity between the source and target domains and then transfer the most similar data to the target domain to generate a hybrid complete training sample for LSTM to improve the prediction performance. To evaluate the effectiveness of the transfer learning–based LSTM, this study implements empirical studies with a real-world data set collected from a stretch of I-15 freeway in Utah. Experimental study results indicate that the transfer learning–based LSTM network could effectively predict the traffic flow conditions with a training sample with missing values.
DOI: 10.1061/JTEPBS.TEENG-7638.

Author keywords: Traffic flow prediction; Transfer learning; Long short-term memory (LSTM) network; Missing data.

Introduction

Traffic congestion mitigation and vehicle emission pollution have become significant problems that need to be solved in modern cities. Recently, with the advancement of data collecting, processing, and computation technologies, data-driven approaches offer the possibility of utilizing data-driven and computation technology to efficiently tackle these problems. Intelligent transportation systems (ITS) aims to apply data-driven computing technology to provide more accurate traffic state prediction (TSP) by using massive data created in cities, which helps travelers plan their trips and allows transportation agencies to take actions to mitigate traffic congestion and therefore reduce air pollution. The acquisition of accurate future traffic information has always been complicated due to the stochastic nature of traffic patterns.

TSP is a method that can predict future traffic information based on historical traffic information (Liu et al. 2020; Lv et al. 2015; Ma et al. 2015; Tian et al. 2018a; Zhang and Ge 2013; Zou et al. 2014), which is an effective way to obtain future traffic information. In practice, historical traffic information is usually collected by various stationary traffic sensors, which can be easily retrieved because it is collected by fixed traffic detectors (e.g., inductive loops and radar detectors) on freeways. However, those collected historical data usually contain missing values and significantly limit traffic state prediction. Fig. 1 illustrates a comparison of long short-term memory (LSTM) predictions using training set with and without missing data. As shown in the figure, different lines represent LSTM predictions with missing data, predictions without missing data, and the ground truth. We can clearly observe that the pattern of LSTM predictions with missing data is significantly different from LSTM predictions without data and ground truth. Hence, there is an urgent need for exploring an efficient countermeasure to address data flawed issues in TSP tasks.

Recently, the effectiveness of transfer learning models in improving the performance of machine learning models in various fields has been recognized, such as environmental science (Chen et al. 2021; Lv et al. 2019), quantum chemistry (Vermeire and Green 2021), bioinformatics (Giorgi and Bader 2018), transportation (Huang et al. 2021; Zhang et al. 2019; Wan et al. 2022; Kasundra et al. 2022), and so on. In the transportation field, people mainly utilized transfer learning model for intercity transfers to solve the insufficient data limitations in target cites (Huang et al. 2021; Wei et al. 2016). However, addressing flawed data problems (e.g., data containing missing or error values) with transfer learning methods has not been well investigated in the literature. Therefore, this research aims to develop an innovative framework, named transfer learning–based LSTM, to solve the missing data problem in TSP tasks.

More specifically, transfer learning is used to find the traffic detector stations with complete monitoring data that are similar to the target domain and then create a hybrid training sample consisting of data from target and source domains. Then, the LSTM network is trained using hybrid training sample. This study makes significant contributions to the literature from the following perspectives: (1) an innovative transfer learning–based LSTM model is proposed for TSP; and (2) an advanced method is provided to tackle the flawed data problem in TSP tasks.

The remainder of this paper is organized as follows. "Literature Review" part provides the review of existing studies related to missing data problems, transfer learning, and traffic state prediction.

[1]Postdoctoral Researcher, Dept. of Civil and Environmental Engineering, Univ. of Utah, 110 S. Central Campus Dr., Suite 2000, Salt Lake City, UT 84112 (corresponding author). ORCID: https://orcid.org/0000-0002-6895-0921. Email: zhao.zhang@utah.edu

[2]Assistant Professor, Dept. of Civil Engineering, McMaster Univ., 1280 Main St. West, John Hodgins Engineering Bldg., Hamilton, ON, Canada L8S 4L7. ORCID: https://orcid.org/0000-0002-1739-5788. Email: haoyang@mcmaster.ca

[3]Assistant Professor, Dept. of Civil and Environmental Engineering, Univ. of Maryland, 1173 Glenn L. Martin Hall, College Park, MD 20742. Email: xtyang@umd.edu

Note. This manuscript was submitted on July 31, 2022; approved on June 5, 2023; published online on July 18, 2023. Discussion period open until December 18, 2023; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Transportation Engineering, Part A: Systems***, © ASCE, ISSN 2473-2907.

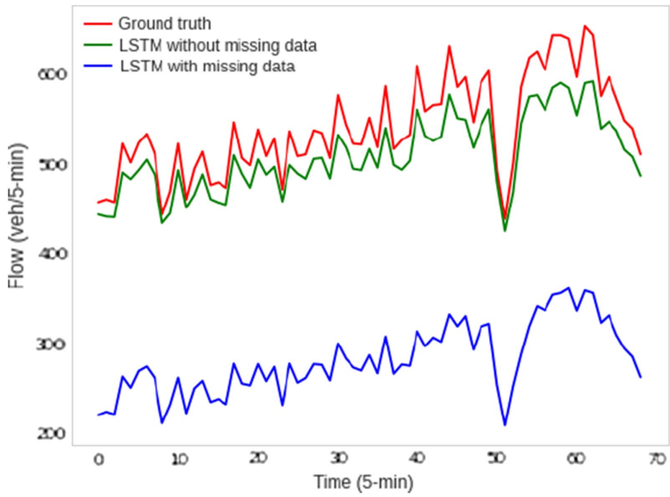

Fig. 1. Comparison LSTM performance between with and without missing data.

The transfer learning–based LSTM framework is presented in the "Methodology" part. The "Case Study Results" are presented in the experimental study part on real-world data from US interstate freeway I-15. The last section of "Conclusions and Future Research Directions" summarizes the key findings and future research directions.

Literature Review

Missing Data Problem

Missing data can be caused by many reasons in collected data of traffic flow, such as malfunction of the sensor, manual system closure, and errors in signal transmission (Tian et al. 2018a). Therefore, missing data imputation is a hot topic, and many methods have been developed for reducing the impact of missing data. Those methods can be generally classified into three categories. The first category is interpolation, including temporal-neighboring and pattern-similar imputation methods (Tak et al. 2016). However, interpolation models cannot make full use of local daily flow variation information to improve model performance (Zhong et al. 2005). The second category is statistical learning methods, such as Markov chain Monte Carlo (MCMC) (Ni and Leonard 2005) and probabilistic principal component analysis (PPCA) (Qu et al. 2009; Tipping and Bishop 1999). It can obtain traffic flow information by using the statistical characteristics of traffic flow, but the accuracy is low because these approaches are based on prior knowledge.

The third category is data-driven prediction approaches, including autoregressive integrated moving average (ARIMA) (Zhong et al. 2004), Bayesian networks (Ghosh et al. 2007; Zhang et al. 2004), neural networks (Dia 2001; Vlahogianni et al. 2005), support vector regression (Castro-Neto et al. 2009), and LSTM networks (Tian et al. 2018a). With the advancement of data collecting, processing, and computation technologies, data-driven approaches are more popular and efficient to solve the missing data problems in traffic flow prediction recently.

Transfer Learning

For traditional machine learning models, the input feature space and data distribution patterns are the same because the basic assumption is the training data and testing data are collected from the same domain. However, this assumption does not hold in many real-world applications because the training data are expensive or difficult to obtain (Weiss et al. 2016). Hence, it is important to create high-performance learners trained with similar and more easily obtained data from different domains. This methodology is referred to as transfer learning (Weiss et al. 2016). Transfer learning is a new branch of artificial intelligence (AI) that has gained increasing interest because it could allow the patterns of training samples to be different (Lv et al. 2019).

More specifically, transfer learning is an innovative method that uses well-established knowledge from a similar source domain to improve the learning efficiency in the target domain (Weiss et al. 2016). Hence, it has been widely implemented in various fields to solve the problem of limited data in target domain to train the model. In real-world applications, unacceptable modeling accuracy tends to be caused by unavailability of training data or training data containing missing values at target stationary detectors. In this research, transfer learning is utilized to transfer knowledge from the complete data sequences obtained at nearby detectors or detectors located on similar roadway geometry to an incomplete sequence in the target domain.

Traffic State Prediction

During the last decades, many data-driven models have been developed to predict the short-term traffic state. Those models can be generally grouped by parametric methods and nonparametric methods (Yu et al. 2021). Parametric methods mainly include ARIMA models (Shi et al. 2014; Van Der Voort et al. 1996; Williams and Hoel 2003) and the Kalman filter (Guo et al. 2014; Ojeda et al. 2013). They cannot obtain a satisfying performance under irregular traffic variations. To solve this limitation, nonparametric methods are developed to obtain the acquisition of nonlinear laws from historical data. These methods mainly include k-nearest neighbors (Cai et al. 2016; Dell'acqua et al. 2015; Sun et al. 2018; Wu et al. 2014), Bayesian model (Wang et al. 2014; Xu et al. 2014), support vector machine (SVM) (Cai et al. 2016; Castro-Neto et al. 2009), and artificial neural network (ANN) (Chen 2017; Smith and Demetsky 1994). But the performance of nonparametric methods is heavily dependent on data quality and quantity of training data.

Recently, various deep learning models have been extensively used in TSP tasks to improve modeling accuracy. In comparison with other deep learning networks, the recurrent neural network (RNN) could better capture the temporal evolution of traffic flow by self-loops and chainlike structures (Qu et al. 2022). But traditional RNN models have following limitations (Gers et al. 1999; Ma et al. 2015): (1) traditional RNNs cannot train time series with long time lags, and (2) traditional RNNs rely on predetermined time lags to learn the temporal sequence processing, but it is challenging to find the optimal time window size in an automatic way. Hochreiter and Schmidhuber (1997) proposed a LSTM network that is a special RNN architecture. The aforementioned constraints of traditional RNNs could be solved by LSTM because it can learn information with long time spans and determine the optimal time lags in an automated manner (Ma et al. 2015; Yu et al. 2021). These advantages made the LSTM network extensively deployed for traffic state prediction (Do et al. 2019; Fu et al. 2016; Kang et al. 2017; Luo et al. 2019; Ma et al. 2015; Mackenzie et al. 2019; Tian et al. 2018a, b; Yang et al. 2019). In this paper, the LSTM network is utilized for traffic flow prediction with transferred historical traffic flow data.

In summary, a transfer learning–based LSTM framework that can transfer knowledge from nearby detectors or detectors with similar roadway geometry to overcome the missing data problem in traffic flow prediction is still lacking. This paper focuses on

J. Transp. Eng., Part A: Systems, 2023, 149(10): 04023095

filling the gap by proposing a transfer learning–based LSTM model for TSP with missing data.

Methodology

To deal with the training sample with missing data problems in traffic flow forecasting, the transfer learning–based LSTM model is constructed. As shown in Fig. 2, the proposed transfer learning–based LSTM includes four key steps. Firstly, the experiment data containing an incomplete data sequence in the target domain are obtained from a detector sensor located on a freeway traffic monitoring station. Secondly, a dynamic time warping (DTW) algorithm is utilized to pinpoint the traffic detector stations with complete traffic flow data that are similar to the target domain. The traffic flow data from these traffic detector stations are named as the source domain. Thirdly, hybrid data from target and source domains are used to construct the training sample. Finally, the LSTM network is trained with hybrid data and the trained model is used to predict the future traffic flow. The transfer learning and LSTM are described in the following sections.

Transfer Learning

Transfer learning is one of the key methods utilized in this study. The target and source domains are two key components of learning, and the essential basis of transfer learning is to identify the similarity between source and target domains. The source domain will have a negative impact on machine learning model performance for the target domain if the pattern of the source domain is different from the target domain. It is important to avoid negative transfer and select an appropriate criterion to measure similarity between target domain and source domains (Pan and Yang 2009). According to the existing literature (Folgado et al. 2018; Fu 2011; Li et al. 2020), there are several similarity measurement criteria that have widely implemented, including Euclidean distance (ED), Kullback-Leibler divergence (K-L divergence), Pearson correlation coefficient (Pearson), longest common subsequence (LCSS), and DTW, among others.

In general, the inherent data properties of the time series are the key to selecting a proper similarity measurement criterion (Chen et al. 2021). The ED method may lead to the inappropriate transmission of time information because it requires the length of the time series to be equal. The main drawback of K-L divergence is that distance and asymmetry are not considered. Pearson's correlation coefficient is good at measuring the strength of the correlation between two variables, but it is an ineffective method for dealing with nonlinear scenarios. LCSS mainly works for shape similarity rather than spatial similarity and is very time-consuming. The DTW algorithm could accurately measure the similarity between the patterns of two time-series data because it can use time series of different lengths (Chen et al. 2021). Hence, DTW was utilized as the fundamental similarity measurement criterion in this research considering the length of continuous gaps in traffic flow data varies.

DTW can be used to measure the similarity between two time-series data and select the shortest distance between values because it is a nonlinear programming technique that involves model similarity matching for time series by bending and aligning the time axis (Chen et al. 2021). In this study, incomplete time-series traffic flow data, named as the target domain data, are defined as D_T. This series can be denoted as $D_T = \{D_T^1, D_T^2, \ldots, D_T^m\}$, which is incomplete time-series traffic flow data with missing values, where m is the sample size of target domain. The important step is to select source domain time-series data D_S, where $1 \leq S \leq N$, $S \neq T$, and N is the total number of traffic detector stations. The detailed procedures of the DTW algorithm are described in Table 1, which is utilized to determine the best source domain for the target domain according to the following formula:

$$D_n = \min_{D_S} \mathrm{DTW}(D_S, D_T), \quad 1 \leq S \leq N \text{ and } S \neq T \tag{1}$$

where D_n = appropriate source domain among all domains of traffic detector stations.

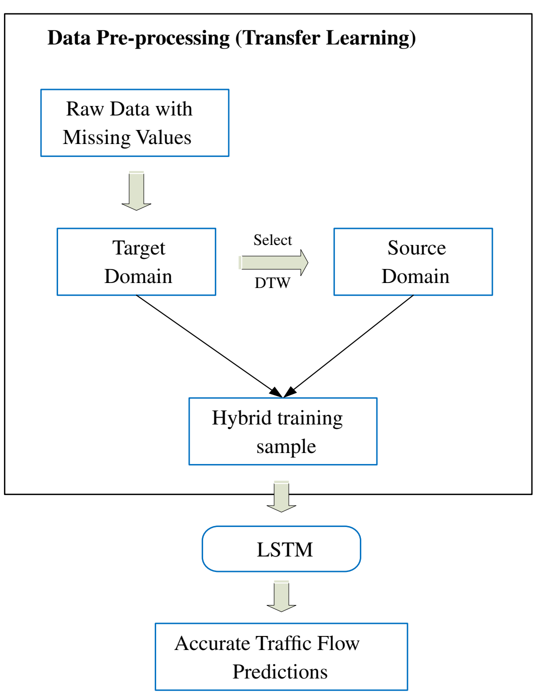

Fig. 2. Diagram of transfer learning–based LSTM with missing data.

Long Short-Term Memory Network

In this study, LSTM is used to generate accurate traffic flow predictions with transferred training samples. LSTM is a special RNN architecture; the LSTM network was proposed by Hochreiter and Schmidhuber (1997), which could overcome the vanishing gradient problem of traditional RNN. A typical LSTM network is composed of one input layer, one recurrent hidden layer (memory block), and one output layer. The memory block contains memory cells with self-connections memorizing the temporal state, and pair of adaptive, multiplicative gating units to control information flow in the block. The typical architecture of an LSTM network is illustrated in Fig. 3.

The input of the historical traffic flow sequence is denoted as $x = (x_1, x_2, \ldots, x_T)$ (where T is the prediction period), and the output sequence is $y = (y_1, y_2, \ldots, y_T)$ can be iteratively calculated as follows:

$$i_t = \mathrm{sig}(w_i[x_t, H_{t-1}] + \eta_i) \tag{2}$$

$$f_t = \mathrm{sig}(w_f[x_t, H_{t-1}] + \eta_f) \tag{3}$$

$$\widetilde{C}_t = \tanh(w_{\widetilde{C}}[x_t, H_{t-1}] + \eta_{\widetilde{C}}) \tag{4}$$

$$o_t = \mathrm{sig}(w_o[x_t, H_{t-1}] + \eta_o) \tag{5}$$

Table 1. Details procedure of the DTW algorithm

Step	Dynamic time warping algorithm
Step 1	Suppose $A = \{a_1, a_2, \ldots, a_n\}$ and $B = \{b_1, b_2, \ldots, b_m\}$ represent two time-series data, n and m are the lengths of two sequences. The two time-series sequences can be formulated as an $n \times m$ distance matrix $\boldsymbol{D}_{\boldsymbol{n}\times\boldsymbol{m}}$, and the elements $d_{ij} = D(a_I, b_j)\|i \in [1,n], j \in [1,m]$, of the distance matrix $\boldsymbol{D}_{\boldsymbol{n}\times\boldsymbol{m}}$ denote the distance between a_i and b_j.
Step 2	$P = \{p_1, p_2, \ldots, p_k\}$ represent the optimal warping path in DTW, which consists of adjacent elements in the matrix $\boldsymbol{D}_{\boldsymbol{n}\times\boldsymbol{m}}$ that represents the kth element of P. The warping path must meet the following conditions: (1) $\max(m,n) \le k \le m + n - 1$. (2) Boundary limits: $p_1 = d_{11}$ and $p_k = d_{nm}$. (3) Two adjacent elements in P must be adjacent in $\boldsymbol{D}_{\boldsymbol{n}\times\boldsymbol{m}}$ and extend forward, namely $P_k = \{\alpha, \beta\}$ and $w_{k+1} = \{\alpha', \beta'\}$. The corresponding points between the two time-series data must not intersect, i.e., $0 \le \alpha' - \alpha \le 1$ and $0 \le \beta' - \beta \le 1$.
Step 3	Calculate the DTW distance $D_{\mathrm{DTW}}(i,j)$. The warping path can be examined by dynamic programming using following formulas: $D_{\mathrm{DTW}}(i,j) = \min\{D_{\mathrm{DTW}}(i-1,j-1), D_{\mathrm{DTW}}(i,j-1), D_{\mathrm{DTW}}(i-1,j)\} + d_{ij} D_{\mathrm{DTW}}(1,1) = d_{11}$
Step 4	The DTW distance for the two time-series data is the $D_{\mathrm{DTW}}(i,j)$ at the endpoint of the two sequences.

Source: Adapted from Chen et al. (2021).

$$C_t = i_t \odot C_{t-1} + f_t \odot \widetilde{C}_t \tag{6}$$

$$H_t = C_t \odot o_t \tag{7}$$

where i_t, f_t, and o_t = input gate, forget gate, and output gate, respectively; $\odot$ = scalar product; w represents the weights matrices; η represents the offset vector; C_{t-1} = state of the previous cell at time t–1; and $\tanh(\cdot)$ = activation function

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{8}$$

Also, $\mathrm{sig}(\cdot)$ denotes the logistic sigmoid function

$$\mathrm{sig}(x) = \frac{1}{1 + e^{-x}} \tag{9}$$

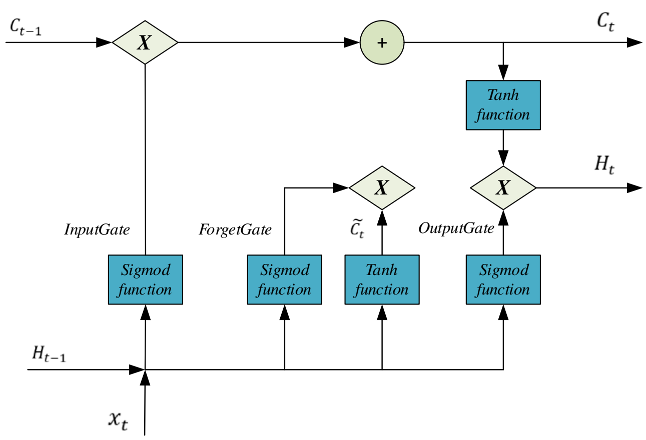

Fig. 3. Structure of LSTM network.

Experimental Study

Experimental Setup

In this study, the Performance Measurement System (PeMS) data collected from interstate freeway I-15 were utilized to validate the effectiveness of the proposed transfer learning–based LSTM network with missing data. PeMS data are the most commonly used data type in traffic flow forecasting tasks (Fu et al. 2016; Qu et al. 2022; Wu et al. 2018; Yu et al. 2021). The studied freeway corridor and detector stations are illustrated in Fig. 4.

In the case study, the separate freeway segment in I-15 has 10 detectors. Station 407 was used for testing the proposed transfer

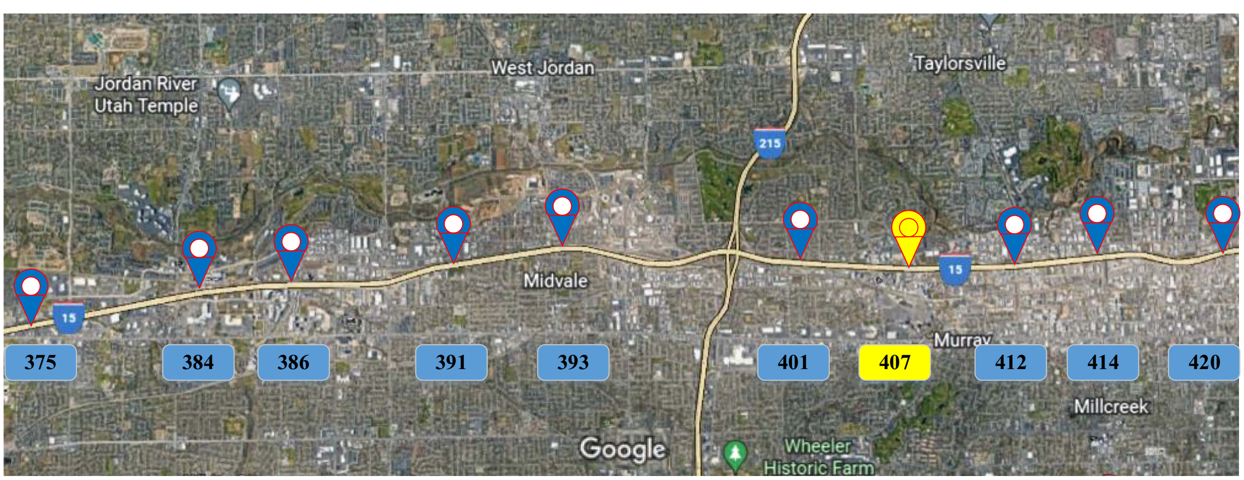

Fig. 4. Deployment of freeway corridor and stations. (Map © 2023 Google.)

learning–based LSTM network with missing data. The other marked detector stations were used for conducting the transfer learning for target Station 407. Traffic flow data were collected from August 2 to August 11, 2021. There were 288 observations per detector per day because the data are collected every 5 min.

To evaluate the performance of the proposed transfer learning–based LSTM network with missing data, testing cases were constructed regarding the basic TSP problem with random missing data. To further test the effectiveness of methods under this condition, we created the missing data scenarios by artificially removing the traffic flow records in the training data to simulate the situation of the device malfunctioning. The robustness analysis was implemented to show the capability of dealing with the unpredictable missing inputs in the training process. Theoretically, the proposed transfer learning–based LSTM is efficient for predicting the traffic flow with missing data on the target locations. To validate this feature, a certain portion of the training data set was randomly removed. The test set is the complete data set. In the robustness study, 30%, 50%, and 70% of the training data were randomly removed, and transfer learning was used to supplement those missing data, and the testing data were kept unchanged.

Table 2 lists all the predetermined parameters of the proposed method in the setup of experiments. Time-series data were expressed in the appropriate format for the LSTM network. Generally, the time-series data are improper to feed into LSTM directly because they consists of several tuples (time, value). Hence, the sliding window technique was used to reconstruct original time-series data. The sliding window technique is presented in Table 3. For example, the traffic flows for the prior three timestamps of the moment $T(i = 1, 2, 3)$ were used as input for the LSTM network, and the traffic flow of moment T is the output. The same data input and output are used for benchmark benchmark methods [e.g., ANN, SVM, andf random forest (RF)].

In this study, the raw flow data were normalized into a range from zero to one by Eq. (10) given the requirements of the machine learning model

Table 2. Prefixed parameters of LSTM network

Parameter	Value
Training set size	2,880
Testing set size	864
Learning rate	0.001
Optimizer	Adam
Activation function	Rectified linear unit (ReLU)
Hidden layer	3
Epochs	1,000

Table 3. Data windowing

Input	Output
$X_{T-i}(i = 1, 2, 3)$	X_T

$$x_n = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{10}$$

where x_n = normalized raw flow data; x_i = flow data; and $x_{\max}$ and $x_{\min}$ = minimum and maximum raw flow data, respectively. Finally, the prediction needs to be denaturable.

Performance Index

To evaluate the accuracy of predictions, this research selects three common prediction evaluation indexes root mean square error (RMSE), mean absolute percentage error (MAPE), and mean absolute error (MAE) (Chen et al. 2018; Qu et al. 2022; Yu et al. 2021) of each dimension as the performance metric, which are defined in Eqs. (11)–(13). The smaller the value of these three evaluation methods indicates better performance of the model

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)2} \tag{11}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i} \times 100\% \tag{12}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{13}$$

where y_i = observed traffic speed and flow; and $\hat{y}_i$ = estimated traffic speed and flow.

Results Analysis

The similarity results between the target domain and source domains are presented in Table 4. As indicated in the table, Station 414 was determined to be the most appropriate source domain for the target domain (Station 407). Then, we used the traffic flow data of Station 414 to transfer to Station 407 and then combined those data to get a hybrid training sample for the LSTM network.

Table 5 summarizes the traffic flow prediction results from LSTM and other benchmark methods (e.g., SVM, RF, and ANN) of Station 407 with missing rates of 0.3, 0.5, and 0.7. For the missing rate of 0.3, LSTM generated a 42.93 vehicles/5-min RMSE, a 9.30% MAPE, and 28.86 vehicles/5-min MAE.

As indicated in the table, in comparison with these benchmark methods, the transfer learning–based LSTM obtained the best modeling performance under the missing rate of 0.3. For the missing rate of 0.5, the transfer learning–based LSTM obtained the most accurate predictions, which yielded a 43.03 vehicles/5-min RMSE, a 9.33% MAPE, and 28.86 vehicles/5-min MAE. For the missing rate of 0.7, the transfer learning–based LSTM can still achieve best prediction results (e.g., a 43.06 vehicles/5-min RMSE, a 9.34% MAPE, and 29.01 vehicles/5-min MAE) compared with benchmark models. All results proved that the proposed transfer

Table 4. DTW correlation ($\times 10^7$) results between the target domain (Station 407) and source domains

Missing rate	375	384	386	391	393	401	412	414	420
0.3	0.139	0.146	0.146	0.123	0.121	0.143	0.137	**0.136**	0.136
0.5	0.174	0.180	0.180	0.185	0.188	0.177	0.169	**0.167**	0.172
0.7	0.173	0.175	0.176	0.190	0.194	0.175	0.169	**0.169**	0.170
Average	0.162	0.167	0.167	0.166	0.168	0.165	0.158	**0.158**	0.159

Note: Bold denotes the best performance.

learning–based LSTM could perform well for traffic flow forecasting with missing data.

Figs. 5(a–c) show the comparison of prediction results of SVM, RF, ANN, and LSTM with ground truth. It can be clearly seen that the line of LSTM could better fit the ground truth under different missing rates compared with benchmark models. This demonstrates the excellent prediction performance of transfer learning–based LSTM under different missing rates.

Table 5. Prediction results of spots on normal segment with missing data

Method	Missing rate	RMSE	MAPE	MAE
SVM	0.3	56.83	16.67	39.54
RF	0.3	52.40	11.78	36.14
ANN	0.3	48.40	13.20	35.85
LSTM	0.3	**42.93**	**9.30**	**28.86**
SVM	0.5	52.57	16.00	38.20
RF	0.5	49.94	11.26	34.50
ANN	0.5	49.87	14.10	37.57
LSTM	0.5	**43.03**	**9.33**	**28.86**
SVM	0.7	53.17	16.05	38.15
RF	0.7	50.25	11.50	35.38
ANN	0.7	49.44	13.68	36.92
LSTM	0.7	**43.06**	**9.34**	**29.01**

Note: Bold denotes the best performance.

To further confirm this finding, the prediction results obtained by the transfer learning–based LSTM were compared with ground truth. In Figs. 6(a–c), the prediction results can be seen as fitting the ground truth well if the coefficient of the trend line is close to one and the intercept is close to zero (Yuan et al. 2021). For the missing rate of 0.3, the coefficient was 0.97 and the intercept was 10.33 for LSTM. For the missing rate of 0.5, the coefficient was 0.97 and the intercept was 10.63 for LSTM. For the missing rate of 0.7, the coefficient was 0.97 and the intercept was 12.82 for LSTM. These results indicate that the transfer learning–based model could achieve stable and accurate prediction performance under different missing rates.

To further verify the robustness of the proposed transfer learning–based LSTM, Fig. 7 illustrates the prediction error of the transfer learning–based LSTM with different missing rates under different traffic dynamic orders. The traffic conditions are divided into three categories according to traffic volume: low traffic volume (0–300 vehicles/5-min), medium traffic volume (300–550 vehicles/5-min), and high traffic volume (>550 vehicles/5-min). As shown

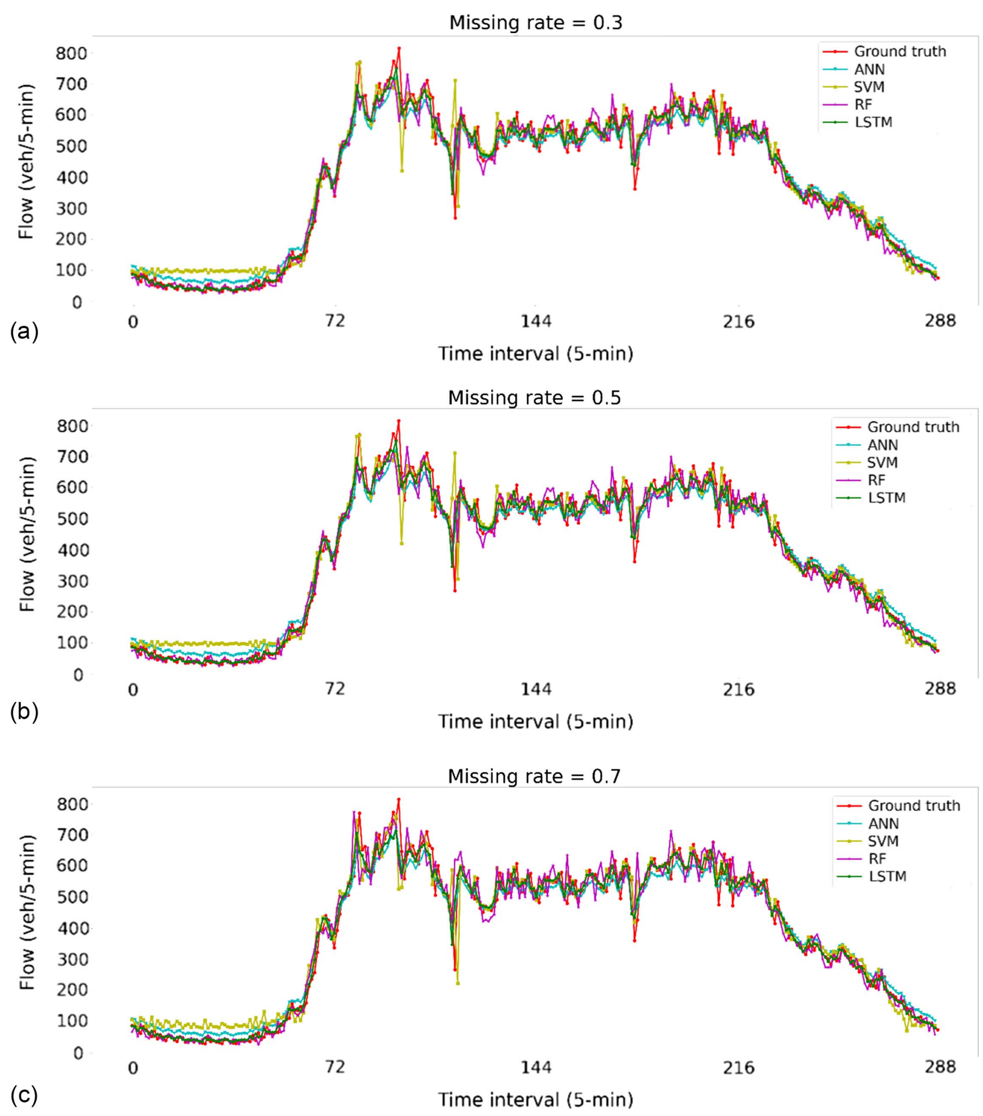

Fig. 5. Prediction results versus ground truth: (a) missing rate = 0.3; (b) missing rate = 0.5; and (c) missing rate = 0.7.

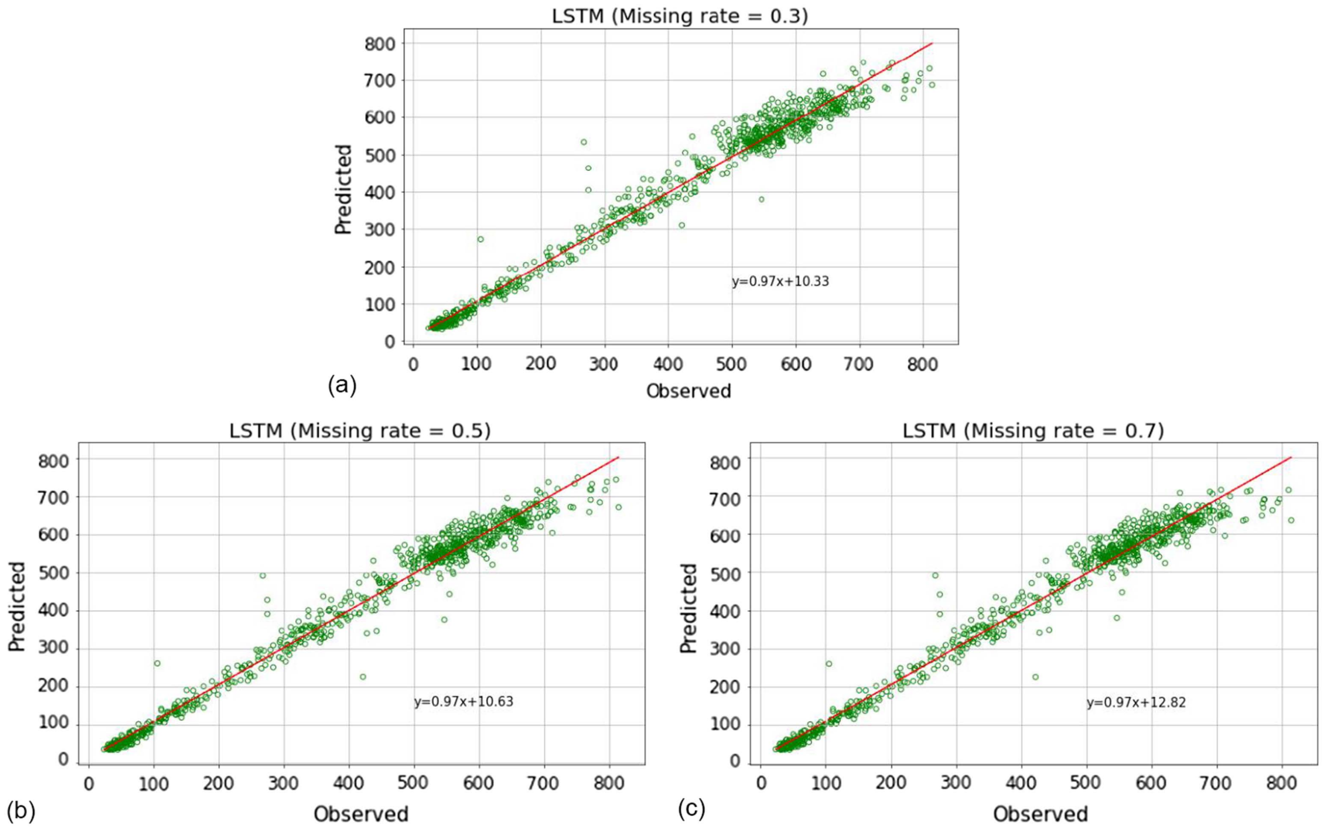

Fig. 6. Predicted flow by LSTM: (a) missing rate = 0.3; (b) missing rate = 0.5; and (c) missing rate = 0.7.

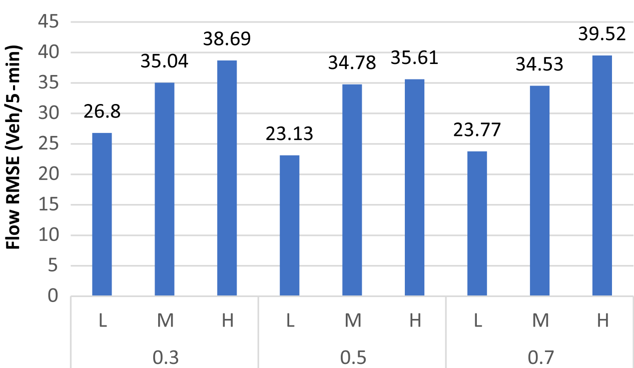

Fig. 7. Prediction performance of transfer learning–based LSTM with different missing rates under different traffic volumes.

in the figure, the performance of transfer learning–based LSTM was downgraded with the traffic volume increasing with different missing rates.

Conclusions and Future Research Directions

Precise traffic flow prediction plays a significant role in the successful operation of ITS on freeways. However, traffic detectors may provide incomplete traffic information, which is a major constraint for existing methods to get precise traffic predictions. To overcome this limitation, this paper introduces an advanced transfer learning–based LSTM network for traffic flow forecasting with incomplete traffic information. This new method could transfer similar data from a source domain to a target domain to generate a hybrid training sample. This attribute makes the transfer learning–based LSTM network overcome the limitation of training samples with missing values and can then improve the prediction performance of LSTM.

Experimental study results indicated that the transfer learning–based LSTM network could effectively predict the traffic flow conditions with training sample with missing values. Hence, the proposed transfer learning–based LSTM network has the potential to help transportation agencies to obtain more accurate traffic flow information to propose more efficient traffic control strategies.

The effectiveness of the transfer learning–based LSTM approach has been proven. However, transfer learning–based LSTM traffic flow forecasting still needs more future studies. More specifically, more advanced machine learning algorithms and similarity evaluation criterion and its application on urban freeway networks are worth studying.

Data Availability Statement

Some or all data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments

This research is supported by the project "CMMI #2047268 CAREER: Physics Regularized Machine Learning Theory: Modeling Stochastic Traffic Flow Patterns for Smart Mobility Systems" funded by the National Science Foundation (NSF).

References

Cai, P., Y. Wang, G. Lu, P. Chen, C. Ding, and J. Sun. 2016. "A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting." *Transp. Res. Part C Emerging Technol.* 62 (Jan): 21–34. https://doi.org/10.1016/j.trc.2015.11.002.

Castro-Neto, M., Y. S. Jeong, M. K. Jeong, and L. D. Han. 2009. "Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions." *Expert Syst. Appl.* 36 (3): 6164–6173. https://doi.org/10.1016/j.eswa.2008.07.069.

Chen, D. 2017. "Research on traffic flow prediction method based on BP neural network." *IEEE Trans. Ind. Inf.* 13 (4): 2000–2008. https://doi.org/10.1109/TII.2017.2682855.

Chen, M., X. Yu, and Y. Liu. 2018. "PCNN: Deep convolutional networks for short-term traffic congestion prediction." *IEEE Trans. Intell. Transp. Syst.* 19 (11): 3550–3559. https://doi.org/10.1109/TITS.2018.2835523.

Chen, Z., H. Xu, P. Jiang, S. Yu, G. Lin, I. Bychkov, A. Hmelnov, G. Ruzhnikov, N. Zhu, and Z. Liu. 2021. "A transfer learning-based LSTM strategy for imputing large-scale consecutive missing data and its application in a water quality prediction system." *J. Hydrol.* 602 (Nov): 126573. https://doi.org/10.1016/j.jhydrol.2021.126573.

Dell'acqua, P., F. Bellotti, R. Berta, and A. De Gloria. 2015. "Time-aware multivariate nearest neighbor regression methods for traffic flow prediction." *IEEE Trans. Intell. Transp. Syst.* 16 (6): 3393–3402. https://doi.org/10.1109/TITS.2015.2453116.

Dia, H. 2001. "An object-oriented neural network approach to short-term traffic forecasting." *Eur. J. Oper. Res.* 131 (2): 253–261. https://doi.org/10.1016/S0377-2217(00)00125-9.

Do, L. N. N., H. L. Vu, B. Q. Vo, Z. Liu, and D. Phung. 2019. "An effective spatial-temporal attention based neural network for traffic flow prediction." *Transp. Res. Part C Emerging Technol.* 108 (Nov): 12–28. https://doi.org/10.1016/j.trc.2019.09.008.

Folgado, D., M. Barandas, R. Matias, R. Martins, M. Carvalho, and H. Gamboa. 2018. "Time alignment measurement for time series." *Pattern Recognit.* 81 (Sep): 268–279. https://doi.org/10.1016/j.patcog.2018.04.003.

Fu, R., Z. Zhang, and L. Li. 2016. "Using LSTM and GRU neural network methods for traffic flow prediction." In *Proc., 31st Youth Academic Annual Conf. of Chinese Association of Automation (YAC)*, 324–328. New York: IEEE.

Fu, T. C. 2011. "A review on time series data mining." *Eng. Appl. Artif. Intell.* 24 (1): 164–181. https://doi.org/10.1016/j.engappai.2010.09.007.

Gers, F. A., J. Schmidhuber, and F. Cummins. 1999. "Learning to forget: Continual prediction with LSTM." *Neural Comput.* 12 (10): 2451–2471. https://doi.org/10.1162/089976600300015015.

Ghosh, B., B. Basu, and M. O'Mahony. 2007. "Bayesian time-series model for short-term traffic flow forecasting." *J. Transp. Eng.* 133 (3): 180–189. https://doi.org/10.1061/(ASCE)0733-947X(2007)133:3(180).

Giorgi, J. M., and G. D. Bader. 2018. "Transfer learning for biomedical named entity recognition with neural networks." *Bioinformatics* 34 (23): 4087–4094. https://doi.org/10.1093/bioinformatics/bty449.

Guo, J., W. Huang, and B. M. Williams. 2014. "Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification." *Transp. Res. Part C Emerging Technol.* 43 (Jun): 50–64. https://doi.org/10.1016/j.trc.2014.02.006.

Hochreiter, S., and J. Schmidhuber. 1997. "Long short term memory." *Neural Comput.* 9 (8): 1735. https://doi.org/10.1162/neco.1997.9.8.1735.

Huang, Y., X. Song, S. Zhang, and J. J. Q. Yu. 2021. "Transfer learning in traffic prediction with graph neural networks." In *Proc., IEEE Int. Intelligent Transportation Systems Conf. (ITSC)*, 3732–3737. New York: IEEE.

Kang, D., Y. Lv, and Y. Chen. 2017. "Short-term traffic flow prediction with LSTM recurrent neural network." In *Proc., 20th Int. Conf. on Intelligent Transportation Systems (ITSC)*, 1–6. New York: IEEE.

Kasundra, K., V. Garikapati, C. Hoehne, Y. Hou, and S. Young. 2022. *Spatial transferability of machine learning based volume estimation models*. Rep. No. NREL/PO-5400-78829. Golden, CO: National Renewable Energy Lab.

Li, H., J. Liu, Z. Yang, R. W. Liu, K. Wu, and Y. Wan. 2020. "Adaptively constrained dynamic time warping for time series classification and clustering." *Inf. Sci.* 534 (Sep): 97–116. https://doi.org/10.1016/j.ins.2020.04.009.

Liu, Y., J. J. Q. Yu, J. Kang, D. Niyato, and S. Zhang. 2020. "Privacy-preserving traffic flow prediction: A federated learning approach." *IEEE Internet Things J.* 7 (8): 7751–7763. https://doi.org/10.1109/JIOT.2020.2991401.

Luo, X., D. Li, Y. Yang, and S. Zhang. 2019. "Spatiotemporal traffic flow prediction with KNN and LSTM." *J. Adv. Transp* 2019 (Feb): 10. https://doi.org/10.1155/2019/4145353.

Lv, M., Y. Li, L. Chen, and T. Chen. 2019. "Air quality estimation by exploiting terrain features and multi-view transfer semi-supervised regression." *Inf. Sci.* 483 (May): 82–95. https://doi.org/10.1016/j.ins.2019.01.038.

Lv, Y., Y. Duan, W. Kang, Z. Li, and F. Y. Wang. 2015. "Traffic flow prediction with big data: A deep learning approach." *IEEE Trans. Intell. Transp. Syst.* 16 (2): 865–873. https://doi.org/10.1109/TITS.2014.2345663.

Ma, X., Z. Tao, Y. Wang, H. Yu, and Y. Wang. 2015. "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data." *Transp. Res. Part C Emerging Technol.* 54 (May): 187–197. https://doi.org/10.1016/j.trc.2015.03.014.

Mackenzie, J., J. F. Roddick, and R. Zito. 2019. "An evaluation of HTM and LSTM for short-term arterial traffic flow prediction." *IEEE Trans. Intell. Transp. Syst.* 20 (5): 1847–1857. https://doi.org/10.1109/TITS.2018.2843349.

Ni, D., and J. D. Leonard. 2005. "Markov chain Monte Carlo multiple imputation using Bayesian networks for incomplete intelligent transportation systems data." *Transp. Res. Rec.* 1935 (1): 57–67. https://doi.org/10.1177/0361198105193500107.

Ojeda, L. L., A. Y. Kibangou, and C. C. De Wit. 2013. "Adaptive Kalman filtering for multi-step ahead traffic flow prediction." In *Proc., American Control Conf.*, 4724–4729. New York: IEEE.

Pan, S. J., and Q. Yang. 2009. "A survey on transfer learning." *IEEE Trans. Knowl. Data Eng.* 22 (10): 1345–1359. https://doi.org/10.1109/TKDE.2009.191.

Qu, L., L. Li, Y. Zhang, and J. Hu. 2009. "PPCA-based missing data imputation for traffic flow volume: A systematical approach." *IEEE Trans. Intell. Transp. Syst.* 10 (3): 512–522. https://doi.org/10.1109/TITS.2009.2026312.

Qu, Z., H. Li, Z. Li, and T. Zhong. 2022. "Short-term traffic flow forecasting method with M-B-LSTM hybrid network." *IEEE Trans. Intell. Transp. Syst.* 23 (1): 225–235. https://doi.org/10.1109/TITS.2020.3009725.

Shi, G., J. Guo, W. Huang, and B. M. Williams. 2014. "Modeling seasonal heteroscedasticity in vehicular traffic condition series using a seasonal adjustment approach." *J. Transp. Eng.* 140 (5): 04014012. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000656.

Smith, B., and M. J. Demetsky. 1994. "Short-term traffic flow prediction models-A comparison of neural network and nonparametric regression approaches." In *Proc., IEEE Int. Conf. on Systems, Man and Cybernetics*. New York: IEEE.

Sun, B., W. Cheng, P. Goswami, and G. Bai. 2018. "Short-term traffic forecasting using self-adjusting k-nearest neighbours." *IET Intell. Transp. Syst.* 12 (1): 41–48. https://doi.org/10.1049/iet-its.2016.0263.

Tak, S., S. Woo, and H. Yeo. 2016. "Data-driven imputation method for traffic data in sectional units of road links." *IEEE Trans. Intell. Transp. Syst.* 17 (6): 1762–1771. https://doi.org/10.1109/TITS.2016.2530312.

Tian, Y., K. Zhang, J. Li, X. Lin, and B. Yang. 2018a. "LSTM-based traffic flow prediction with missing data." *Neurocomputing* 318 (Nov): 297–305. https://doi.org/10.1016/j.neucom.2018.08.067.

Tian, Y., K. Zhang, J. Li, X. Lin, and B. Yang. 2018b. "Neurocomputing LSTM-based traffic flow prediction with missing data." *Neurocomputing* 318 (Nov): 297–305. https://doi.org/10.1016/j.neucom.2018.08.067.

Tipping, M. E., and C. M. Bishop. 1999. "Mixtures of probabilistic principal component analyzers." *Neural Comput.* 11 (2): 443–482. https://doi.org/10.1162/089976699300016728.

Van Der Voort, M., M. Dougherty, and S. Watson. 1996. "Combining Kohonen maps with ARIMA time series models to forecast traffic flow."

Transp. Res. Part C Emerging Technol. 4 (5): 307–318. https://doi.org/10.1016/S0968-090X(97)82903-8.

Vermeire, F. H., and W. H. Green. 2021. "Transfer learning for solvation free energies: From quantum chemistry to experiments." *Chem. Eng. J.* 418 (Aug): 129307. https://doi.org/10.1016/j.cej.2021.129307.

Vlahogianni, E. I., M. G. Karlaftis, and J. C. Golias. 2005. "Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach." *Transp. Res. Part C Emerging Technol.* 13 (3): 211–234. https://doi.org/10.1016/j.trc.2005.04.007.

Wan, X., H. Liu, H. Xu, and X. Zhang. 2022. "Network traffic prediction based on LSTM and transfer learning." *IEEE Access* 10 (Aug): 86181–86190. https://doi.org/10.1109/ACCESS.2022.3199372.

Wang, J., W. Deng, and Y. Guo. 2014. "New Bayesian combination method for short-term traffic flow forecasting." *Transp. Res. Part C Emerging Technol.* 43 (Jun): 79–94. https://doi.org/10.1016/j.trc.2014.02.005.

Wei, Y., Y. Zheng, and Q. Yang. 2016. "Transfer knowledge between cities." In *Proc., 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, 1905–1914. New York: Association for Computing Machinery.

Weiss, K., T. M. Khoshgoftaar, and D. Wang. 2016. "A survey of transfer learning." *J. Big Data* 3 (1): 1–40. https://doi.org/10.1186/s40537-016-0043-6.

Williams, B. M., and L. A. Hoel. 2003. "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results." *J. Transp. Eng.* 129 (6): 664–672. https://doi.org/10.1061/(ASCE)0733-947X(2003)129:6(664).

Wu, S., Z. Yang, X. Zhu, and B. Yu. 2014. "Improved k-NN for short-term traffic forecasting using temporal and spatial information." *J. Transp. Eng.* 140 (7): 04014026. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000672.

Wu, Y., H. Tan, L. Qin, B. Ran, and Z. Jiang. 2018. "A hybrid deep learning based traffic flow prediction method and its understanding." *Transp. Res. Part C: Emerging Technol.* 90 (May): 166–180. https://doi.org/10.1016/j.trc.2018.03.001.

Xu, Y., Q. J. Kong, R. Klette, and Y. Liu. 2014. "Accurate and interpretable Bayesian MARS for traffic flow prediction." *IEEE Trans. Intell. Transp. Syst.* 15 (6): 2457–2469. https://doi.org/10.1109/TITS.2014.2315794.

Yang, B., S. Sun, J. Li, X. Lin, and Y. Tian. 2019. "Traffic flow prediction using LSTM with feature enhancement." *Neurocomputing* 332 (Mar): 320–327. https://doi.org/10.1016/j.neucom.2018.12.016.

Yu, J., H. Wei, H. Guo, and Y. Cai. 2021. "Urban traffic state prediction based on SA-LSTM." *IOP Conf. Ser. Earth Environ. Sci.* 783 (1): 012153. https://doi.org/10.1088/1755-1315/783/1/012153.

Yuan, Y., Z. Zhang, X. T. Yang, and S. Zhe. 2021. "Macroscopic traffic flow modeling with physics regularized Gaussian process: A new insight into machine learning applications in transportation." *Transp. Res. Part B Methodol.* 146 (Apr): 88–110. https://doi.org/10.1016/j.trb.2021.02.007.

Zhang, C., S. Sun, and G. Yu. 2004. "A Bayesian network approach to time series forecasting of short-term traffic flows." In *Proc., 7th Int. IEEE Conf. on Intelligent Transportation Systems*, 216–221. New York: IEEE.

Zhang, C., H. Zhang, J. Qiao, D. Yuan, and M. Zhang. 2019. "Deep transfer learning for intelligent cellular traffic prediction based on cross-domain big data." *IEEE J. Sel. Areas Commun.* 37 (6): 1389–1401. https://doi.org/10.1109/JSAC.2019.2904363.

Zhang, Y., and H. Ge. 2013. "Freeway travel time prediction using Takagi-Sugeno-Kang fuzzy neural network." *Comput.-Aided Civ. Infrastruct. Eng.* 28 (8): 594–603. https://doi.org/10.1111/mice.12014.

Zhong, M., S. Sharma, and P. Lingras. 2004. "Genetically designed models for accurate imputation of missing traffic counts." *Transp. Res. Rec.* 1879 (1): 71–79. https://doi.org/10.3141/1879-09.

Zhong, M., S. Sharma, and Z. Liu. 2005. "Assessing robustness of imputation models based on data from different jurisdictions examples of Alberta and Saskatchewan, Canada." *Transp. Res. Rec.* 1917 (1): 116–126. https://doi.org/10.1177/0361198105191700114.

Zou, Y., X. Zhu, Y. Zhang, and X. Zeng. 2014. "A space-time diurnal method for short-term freeway travel time prediction." *Transp. Res. Part C Emerging Technol.* 43 (Jun): 33–49. https://doi.org/10.1016/j.trc.2013.10.007.