

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

Zhao Zhang

Ph.D., Postdoctoral Researcher

Department of Civil and Environmental Engineering, University of Utah

100 Central Campus Dr., Suite 2000, Salt Lake City, Utah 84112

Email: zhao.zhang@utah.edu

Xianfeng (Terry) Yang

Ph.D., Assistant Professor

Department of Civil and Environmental Engineering, University of Maryland

1173 Glenn L. Martin Hall, College Park, Maryland 20742

Email: xyang@umd.edu

Word Count: 5900words + 5 Tables * 250 = 7150 words

Submitted to 102nd Transportation Research Board Annual Meeting for Presentation and Publication

Submission Date: August 1, 2022

ABSTRACT

Traffic flow prediction task plays 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 dataset 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.

Keywords: Traffic flow prediction, Transfer learning, LSTM, Missing data

1. INTRODUCTION

With the rapid development of urbanization, traffic congestion mitigation and environmental pollution treatment 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, 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 (1–6), 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.

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 (7, 8), quantum chemistry (9), Bioinformatics (10), transportation (11, 12), etc. In transportation field, people mainly utilized transfer learning model for inter-city transfer to solve the insufficient data limitations in target cities (11, 13). However, addressing flawed data problems with transfer learning methods has not been well investigated in the literature. Therefore, this research aims to develop an innovative framework, named as transfer learning-based LSTM, to solve the missing data problem in TSP tasks. More specifically, transfer learning is used to find out the monitoring stations with complete monitoring data and similar to the target domain and then create a hybrid training sample consisting of data from target and source domains. Then 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 problem; (2) an advanced method is provided to solve the flawed data problem in TSP tasks.

The remainder of this paper is organized as follows. Section 2 reviews existing studies related to missing data problems, transfer learning, and traffic state prediction. The transfer learning-based LSTM framework are presented in Section 3. Section 4 implements the case study on a real-world data from interstate freeway I-15. The last section summarizes the key findings and future research directions.

2. LITERATURE REVIEW

Missing Data Problem

The missing data can be caused by many reasons in the traffic flow, such as malfunction of the sensor, manual system closure, and errors in signal transmission (3). 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: (a) Interpolation, including temporal-neighboring and pattern-similar imputation methods (14). The interpolation models cannot make full use of local daily flow variation information to improve model

performance (15). (b) Statistical learning methods, such as Markov chain Monte Carlo (MCMC) (16) and probabilistic principal component analysis (PPCA) (17, 18). 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. (c) data-driven prediction approaches, including autoregressive integrated moving average (ARIMA) (19), Bayesian networks (20, 21), neural network (22, 23), support vector regression (24), and long short-term memory (LSTM) network (3). 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 basic assumption is the training data and testing data are taken from the same domain, so the input feature space and data distribution characteristics are the same. However, in some real-world machine learning scenarios, this assumption does not hold. There are cases where training data is expensive or difficult to collect. Therefore, there is a need to create high-performance learners trained with more easily obtained data from different domains. This methodology is referred to as transfer learning (25). Transfer learning is a novel branch of artificial intelligence (AI), which allows the distributions of training samples to be different (7). More specifically, transfer learning is a unique technique that uses well-established knowledge from a related source domain to improve the learning efficiency in the target domain (25). Hence, it has been widely implemented in various fields to solve the problem of limited data in target domain to train the model. In this study, due to the large ratio of missing data issues at target stationary detectors, the unavailability of time series of labeled training data leads to unsatisfactory modeling performance. Therefore, transfer learning is applied to transfer knowledge from the complete sequences obtained at nearby detectors or detectors located on similar roadway geometry to an incomplete sequence in the target domain.

Traffic state prediction

In the past decades, many data-driven models have been developed to predict short-term traffic state. Those models can be generally grouped by parametric methods and nonparametric methods (26). Parametric methods mainly include ARIMA model (27–29) and Kalman filter (30, 31). They cannot obtain satisfying performance under irregular traffic variations. To solve this limitation, nonparametric methods are developed to obtain the acquisition of nonlinear laws from historical data. It mainly includes k-Nearest Neighbors (32–35), Bayesian model (36, 37), SVM (24, 33), and ANN (38, 39). But the performance of parametric 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 to other deep learning networks, the recurrent neural network (RNN) could better capture the temporal evolution of traffic flow by self-loops and chain-like structures (40). But traditional RNN models have following limitations (2, 44): (1) Traditional RNNs cannot train time series with long time lags, and (2) Traditional RNNs rely on the predetermined time lags to learn the temporal sequence processing, but it is challenging to find the optimal time window size in an automatic way. To tackle those issues, LSTM Network is a special Recurrent Neural Network (RNN) architecture, was initially proposed by Hochreiter and Schmidhuber (45). The aforementioned constraints could be solved by LSTM since it can learn information with long time spans and determine the optimal time lags in an automated manner (2, 26). These advantages made the

LSTM network extensively deployed for traffic state prediction (2, 3, 46–52). 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 filling the gap by proposing a transfer learning-based LSTM model for TSP with missing data.

3. 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 Figure 1, the proposed transfer learning-based LSTM includes four key steps. Firstly, an incomplete data sequence is collected from a detector sensor located on a freeway monitoring station and defined as the experiment data in the target domain. Secondly, the time series similar method (e.g., dynamic time warping (DTW)) is utilized to find out the monitoring stations with complete monitoring data and similar to the target domain. The data from these stations are named as the source domain. Thirdly, hybrid data from target and source domains to construct the training sample. Finally, train LSTM network with hybrid data and use the trained model to predict the future traffic flow. The transfer learning and LSTM are described in the following sections.

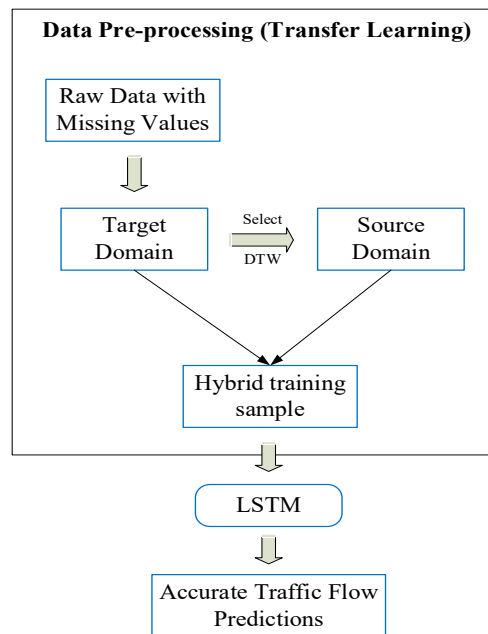


Figure 1 The diagram of transfer learning-based LSTM with missing data

Transfer Learning

Transfer learning is one of the key methods utilized in this study. The target and source domains are two essential components of learning, and the core foundation of transfer learning is to identify the similarity between source and target domains. The source domain data will have a negative impact on machine learning model performance in the target domain if the source domain is weakly related to the target domain. Pan et al. (53) stated that it is important to avoid negative transfer in the transfer learning research and select an appropriate criterion to measure similarity between target domain and source domains. Based on the existing literature (54–56),

the commonly used similarity measurement criteria include Euclidean distance (ED), Kullback-Leibler divergence (K-L divergence), Pearson correlation coefficient (Pearson), longest common subsequence (LCSS), and Dynamic Time Warping (DTW), etc. In general, inherent data properties of the time series is the key to select an appropriate similarity measurement criterion. The ED method may lead to the inappropriate transmission of time information since it requires the length of the time series must be equal. The main disadvantage of K-L divergence is that distance and asymmetry are not considered. Pearson's correlation coefficient is good at dealing with measuring the strength of the correlation between two variables, but it is not ideal method for nonlinear cases. LCSS is mainly works for shape similarity rather than the spatial similarity and is very time consuming. The DTW algorithm can use time series of different lengths and can accurately capture the similarity between the trends of two time series. Hence, considering the length of continuous gaps in traffic flow data varies, DTW is selected as the key similarity measurement criterion in this paper.

DTW can be designed to calculate the similarity between two time series and select the shortest distance between values since it is a nonlinear programming technique that involves model similarity matching for time series by bending and aligning the time axis. 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 $D_T = \{D_T^1, D_T^2, \dots, D_T^m\}$, it is incomplete time series traffic flow data with missing values. 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 detector stations. The detailed procedures of the DTW algorithm are described as Table 1, which is used to determine the best source domain the target domain based on the following constraints:

$$D_n \begin{cases} \min_{D_S} (DTW(D_S, D_T^1)) \\ 1 \leq S \leq N \text{ and } S \neq T \end{cases} \quad (1)$$

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

Table 1 The details procedure of the DTW algorithm

Dynamic Time Warping (DTW) algorithm	
Step 1	Suppose $X = \{x_1, x_2, \dots, x_n\}$ and $Y = \{y_1, y_2, \dots, y_m\}$ 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 $D_{n \times m}$, and the elements $d_{ij} = D(x_i, y_j)$ of the distance matrix $D_{n \times m}$ denote the distance between x_i and y_j .
Step 2	$W = \{w_1, w_2, \dots, w_k\}$ denote the optimal warping path in DTW, which consists of adjacent elements in the matrix $D_{n \times m}$ that represents the K th element of W . The warping path must satisfy the following three conditions: (1) $\max(m, n) \leq k \leq m + n - 1$. (2) Boundary limits: $w_1 = d_{11}$ and $w_k = d_{nm}$. (3) Two adjacent elements in W must be adjacent in $D_{n \times m}$ and extend forward, namely $w_k = \{a, b\}$ and $w_{k+1} = \{a', b'\}$. The corresponding points between the two time series data must not intersect. i.e., $0 \leq a' - a \leq 1$ and $0 \leq b' - b \leq 1$.
Step 3	Calculate the DTW distance $D_{DTW}(i, j)$ between the starting points i and j of the two sequences. The warping path can be examined by dynamic programming using following formulas: $D_{DTW}(i, j) = \min\{D_{DTW}(i-1, j-1), D_{DTW}(i, j-1), D_{DTW}(i-1, j)\} + d_{ij}$ $D_{DTW}(1, 1) = d_{11}$

Step 4	The DTW distance for the two time series data is the $D_{DTW}(i, j)$ at the endpoint of the two sequences.
--------	--

Long Short-Term Memory (LSTM) Network

In this study, Long Short-Term Memory (LSTM) is used to generate accurate traffic flow predictions with transferred training samples. LSTM is a special Recurrent Neural Network (RNN) architecture – LSTM network was proposed by Hochreiter and Schmidhuber (45), 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 Figure 2.

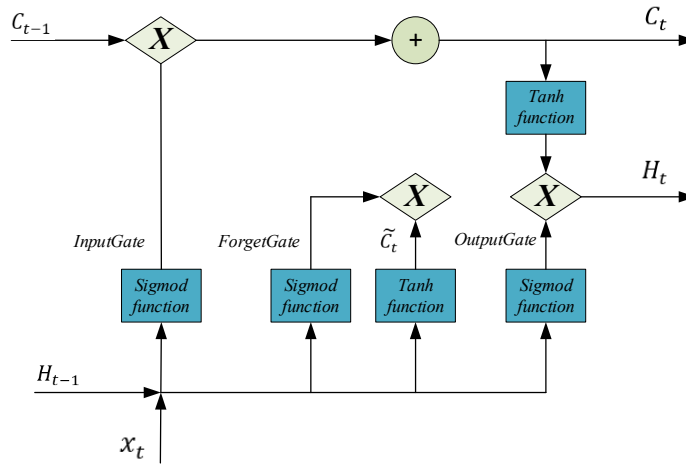


Figure 2 The structure of LSTM network

The current state H_t of an LSTM unit at time t can be expressed by following formula:

$$\text{InputGate} = \text{sig}(w_i[x_t, H_{t-1}] + \eta_i) \quad (2)$$

$$\text{ForgetGate} = \text{sig}(w_f[x_t, H_{t-1}] + \eta_f) \quad (3)$$

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

$$\text{OutputGate} = \text{sig}(w_o[x_t, H_{t-1}] + \eta_o) \quad (5)$$

$$C_t = \text{InputGate} \odot C_{t-1} + \text{ForgetGate} \odot \tilde{C}_t \quad (6)$$

$$H_t = C_t \odot \text{OutputGate} \quad (7)$$

where $\odot$ denotes the scalar product; w denotes the weights matrices; η denotes the offset vector; C_{t-1} denotes the state of the previous cell at time $t-1$; $\tanh(\cdot)$ denotes the activation function:

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

$\text{sig}(\cdot)$ denotes the logistic sigmoid function:

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

4. EXPERIMENTAL STUDY

Experimental Setup

In this study, the Performance Measurement System (PeMS) data collected from interstate freeway I-15 is utilized to validate the effectiveness of the proposed transfer learning-based LSTM network with missing data. PeMS data is the most commonly used data type in traffic flow forecasting tasks (26, 40, 48, 57). The studied freeway corridor and detector stations are illustrated in Figure 3. In the studied scenario, the separate freeway segment in I-15 has 10 detectors. Station 407 marked as yellow is used for testing the proposed transfer learning-based LSTM network with missing data. The detector stations marked as blue are used for conducting the transfer learning for target station 407. Traffic flow data were collected from August 2, 2021, to August 11, 2021. Since the data is collected every 5 minutes, there are 288 observations per detector per day. To evaluate the performance of the proposed transfer learning-based LSTM network with missing data, the testing cases are constructed regarding the basic TSP problem with random missing data. To further test the robustness of methods this case, we investigate the missing data scenarios by artificially removing the traffic flow in the training data to mimic the common device malfunctioning situations. The robustness analysis is conducted to show the capability of dealing with the unpredictable missing inputs in the training phase. Theoretically, the proposed transfer learning-based LSTM is efficient for predicting the traffic flow with missing data on the target locations. To justify this feature, a certain portion of the training dataset is removed. The test set is the complete dataset. In the robustness study, 30%, 50%, and 70% of the training data are removed and transfer learning was used to supplement those missing data, and the testing data keep unchanged.

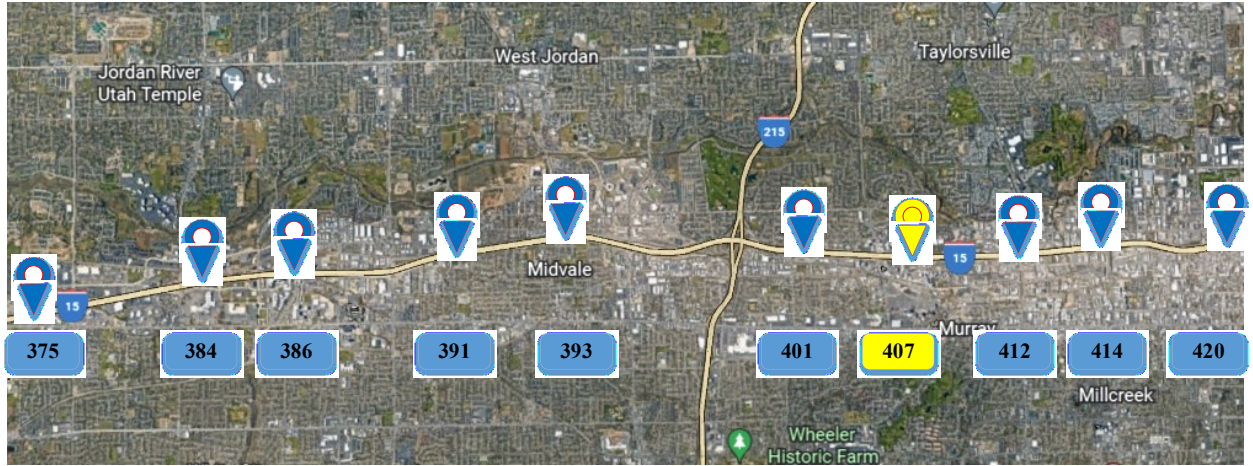


Figure 3 The deployment of freeway corridor and stations

In the setup of the experiments, the prefixed parameters of the proposed method are summarized in Table 2. Time series data need to be expressed in the proper format for the LSTM network. Generally, the time-series data consists of several tuples (time, value), which is inappropriate for feeding them into LSTM. Hence, the sliding window technique is used to reconstruct original time series data. The sliding window technique is presented in Table 3. For example, the traffic flows for the prior 3 timestamps of the moment $T(i = 1, 2, 3)$, are used as input for the LSTM network, and the traffic flow of moment T is the output.

Table 2 The prefixed parameters of LSTM network

Parameter	Value
Training set size	2880
Testing set size	864
The learning rate	0.001
Optimizer	Adam
Activation fuction	Relu
Hidden layer	3
Epoch	1000

Table 3 Data windowing

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

In view of the requirements of machine learning models, the raw flow data were normalized into a range from 0 to 1, using the following equation:

$$x_n = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (10)$$

where, x_n denote the normalized raw flow data; x_i denote the flow data; and x_{max} and x_{min} are the 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 RMSE, MAPE, and MAE (26, 40, 58) of each dimension are used as the performance metric, which are defined in Equations 11 -13. The smaller the value of these three evaluation methods indicate better performance of the model.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y_i^*)^2} \quad (11)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i^* - y_i|}{y_i} * 100\% \quad (12)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - y_i^*| \quad (13)$$

where, y_i is the observed traffic speed and flow and y_i^* is the estimated traffic speed and flow.

Results Analysis

The similarity results between the target domain and source domains are presented in Table 4. As shown in the table, station 414 is 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 407 and then combine those data to get a hybrid training sample for the LSTM network.

Table 4 The DTW correlation ($\times 10^7$) results between the target domain (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

Table 5 summarizes the traffic flow prediction results from LSTM and other benchmark methods (e.g., SVM, RF, and ANN) of the station 407 with missing rates 0.3, 0.5, and 0.7). For the missing rate of 0.3, LSTM generates a 42.93 vehicles/5-minutes of RMSE, a 9.30% of MAPE, and 28.86 vehicles/5-minutes of MAE. As shown in the table, in comparison to these benchmark methods, the transfer learning-based LSTM obtains best modeling performance under the missing rate of 0.3. For the missing rate of 0.5, the transfer learning-based LSTM obtains the most accurate predictions, which yields a 43.03 vehicles/5-minutes of RMSE, a 9.33% of MAPE, and 28.86 vehicles/5-minutes of 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-minutes of RMSE, a 9.34% of MAPE, and 29.01 vehicles/5-minutes of MAE) compared with benchmark models. All results proved that the proposed transfer learning-based LSTM could perform well for traffic flow forecasting with missing data. Figure 4 (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

To further confirm this finding, the prediction results obtained by transfer learning-based LSTM compared with ground truth. In Figure (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. For the missing rate of 0.3, the coefficient is 0.97 and the intercept is 10.33 for LSTM. For the missing rate of 0.5, the coefficient is 0.97 and the intercept is 10.63 for LSTM. For the missing rate of 0.7, the coefficient is 0.97 and the intercept is 12.82 for LSTM. These results

- 1 indicate that the transfer learning-based could achieve stable and accurate prediction
- 2 performance under different missing rates.

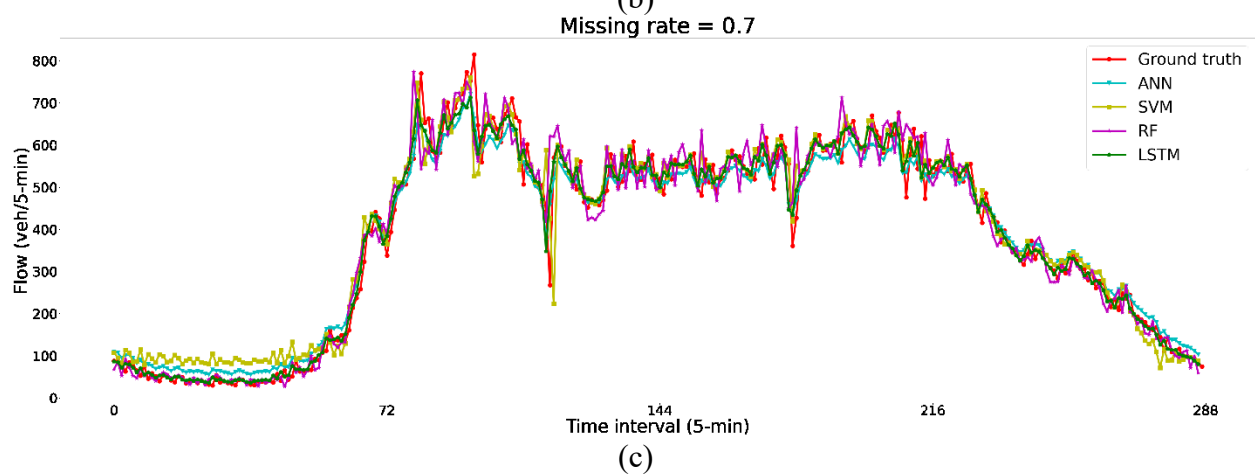
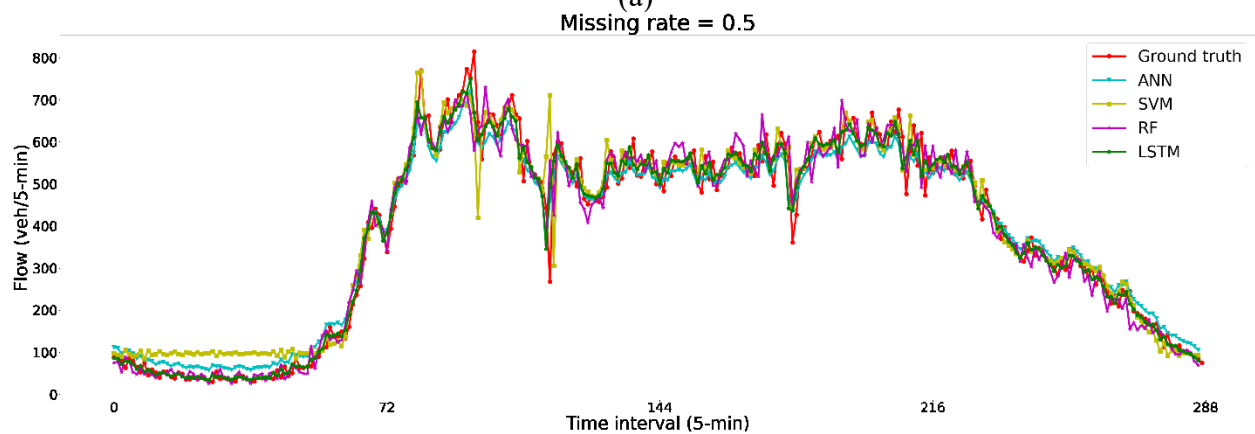
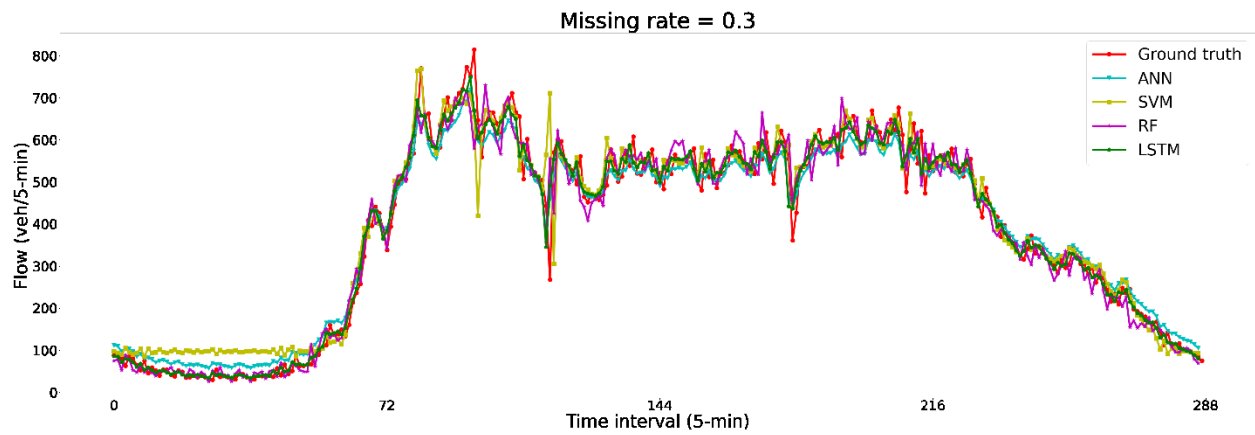


Figure 4 Prediction results versus ground truth

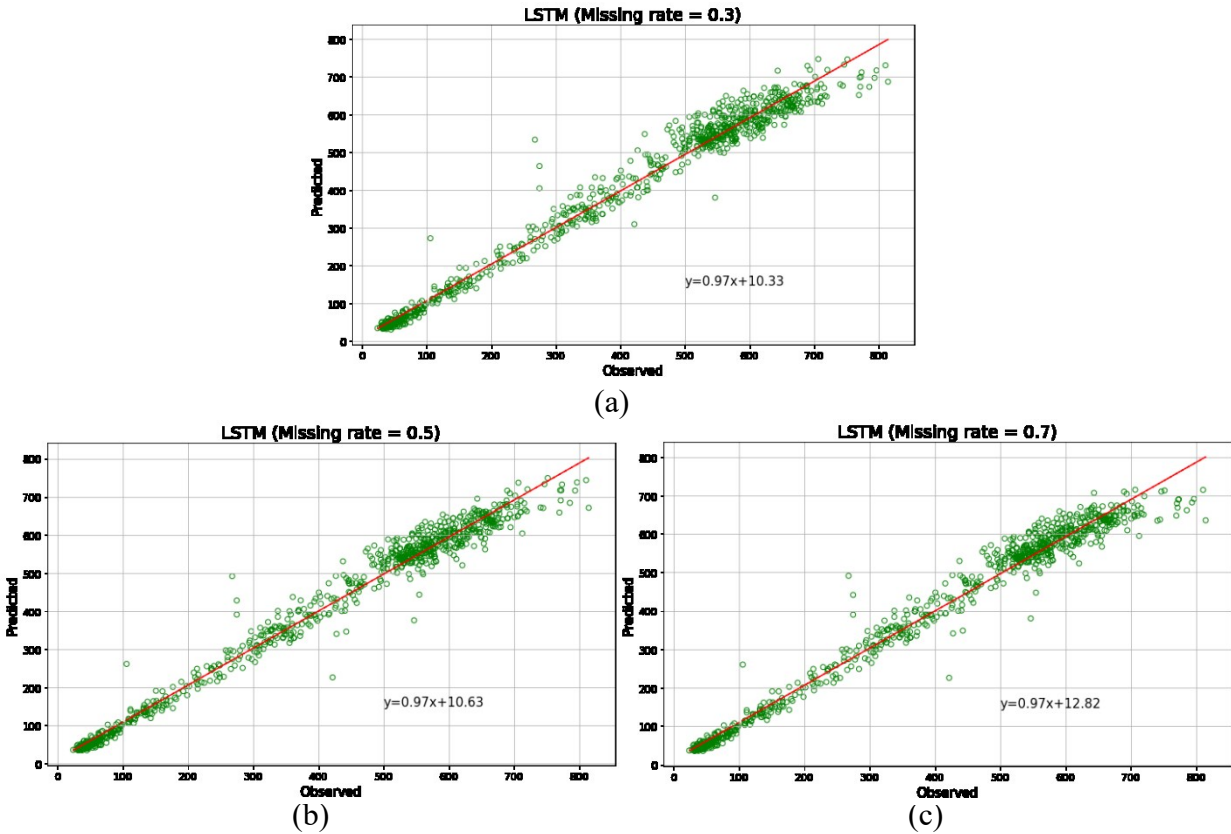


Figure 5 Predicted flow by LSTM

To further verify the robustness of the proposed transfer learning-based LSTM, Figure 6 illustrates the prediction error of 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-minutes), medium traffic volume (300-550 vehicles/5-minutes), and high traffic volume (> 550 vehicles/5-minutes). As shown in the figure, the performance of transfer learning-based LSTM is downgraded with the traffic volume increasing with different missing rates.

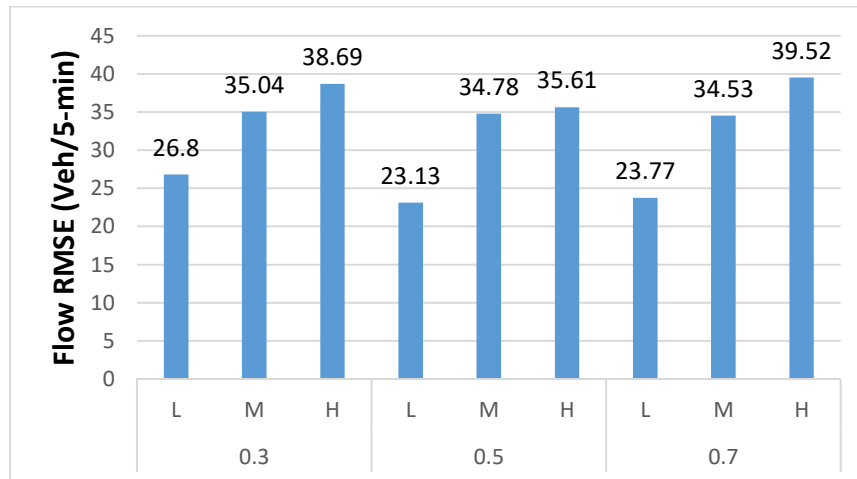


Figure 6 Prediction performance of transfer learning-based LSTM with different missing rates under different traffic volumes

5. CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

Accurate traffic flow prediction plays important 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 source domain to target domain to generate a hybrid training sample. This attribute makes the transfer learning-based LSTM network could overcome the limitation of training sample with missing values and then improve the prediction performance of LSTM. Experimental study results indicate 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 potential to help transportation agencies find better countermeasures to mitigate traffic congestions, improve traffic operation efficiency and safety, and save the freeway operation cost.

The effectiveness of the transfer learning-based LSTM approach has been approved. 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.

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).

AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: study conception and design: X. Yang, Z. Zhang; data collection: Z. Zhang; analysis and interpretation of results: Z. Zhang, X. Yang; draft manuscript preparation: Z. Zhang and X. Yang. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

1. Lv, Y., Y. Duan, W. Kang, Z. Li, and F. Y. Wang. Traffic Flow Prediction with Big Data: A Deep Learning Approach. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 16, No. 2, 2015, pp. 865–873. <https://doi.org/10.1109/TITS.2014.2345663>.
2. Ma, X., Z. Tao, Y. Wang, H. Yu, and Y. Wang. Long Short-Term Memory Neural Network for Traffic Speed Prediction Using Remote Microwave Sensor Data. *Transportation Research Part C: Emerging Technologies*, Vol. 54, 2015, pp. 187–197. <https://doi.org/10.1016/j.trc.2015.03.014>.
3. Tian, Y., K. Zhang, J. Li, X. Lin, and B. Yang. LSTM-Based Traffic Flow Prediction with Missing Data. *Neurocomputing*, Vol. 318, 2018, pp. 297–305. <https://doi.org/10.1016/j.neucom.2018.08.067>.
4. Zou, Y., X. Zhu, Y. Zhang, and X. Zeng. A Space-Time Diurnal Method for Short-Term Freeway Travel Time Prediction. *Transportation Research Part C: Emerging Technologies*, Vol. 43, 2014, pp. 33–49. <https://doi.org/10.1016/j.trc.2013.10.007>.
5. Zhang, Y., and H. Ge. Freeway Travel Time Prediction Using Takagi-Sugeno-Kang Fuzzy Neural Network. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 28, No. 8, 2013, pp. 594–603. <https://doi.org/10.1111/mice.12014>.
6. Liu, Y., J. J. Q. Yu, J. Kang, D. Niyato, and S. Zhang. Privacy-Preserving Traffic Flow Prediction: A Federated Learning Approach. *IEEE Internet of Things Journal*, Vol. 7, No. 8, 2020, pp. 7751–7763. <https://doi.org/10.1109/IIOT.2020.2991401>.
7. Lv, M., Y. Li, L. Chen, and T. Chen. Air Quality Estimation by Exploiting Terrain Features and Multi-View Transfer Semi-Supervised Regression. *Information Sciences*, Vol. 483, 2019, pp. 82–95. <https://doi.org/10.1016/j.ins.2019.01.038>.
8. Chen, Z., H. Xu, P. Jiang, S. Yu, G. Lin, I. Bychkov, A. Hmelnov, G. Ruzhnikov, N. Zhu, and Z. Liu. A Transfer Learning-Based LSTM Strategy for Imputing Large-Scale Consecutive Missing Data and Its Application in a Water Quality Prediction System. *Journal of Hydrology*, Vol. 602, No. June, 2021, p. 126573. <https://doi.org/10.1016/j.jhydrol.2021.126573>.
9. Vermeire, F. H., and W. H. Green. Transfer Learning for Solvation Free Energies: From Quantum Chemistry to Experiments. *Chemical Engineering Journal*, Vol. 418, No. March, 2021. <https://doi.org/10.1016/j.cej.2021.129307>.
10. Giorgi, J. M., and G. D. Bader. Transfer Learning for Biomedical Named Entity Recognition with Neural Networks. *Bioinformatics*, Vol. 34, No. 23, 2018, pp. 4087–4094. <https://doi.org/10.1093/bioinformatics/bty449>.
11. Huang, Y., X. Song, S. Zhang, and J. J. Q. Yu. Transfer Learning in Traffic Prediction with Graph Neural Networks. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, 2021, pp. 3732–3737. <https://doi.org/10.1109/ITSC48978.2021.9564890>.

12. Zhang, C., H. Zhang, J. Qiao, D. Yuan, and M. Zhang. Deep Transfer Learning for Intelligent Cellular Traffic Prediction Based on Cross-Domain Big Data. *IEEE Journal on Selected Areas in Communications*, Vol. 37, No. 6, 2019, pp. 1389–1401. <https://doi.org/10.1109/JSAC.2019.2904363>.
13. Wei, Y., Y. Zheng, and Q. Yang. Transfer Knowledge between Cities. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 1905–1914. <https://doi.org/10.1145/2939672.2939830>.
14. Tak, S., S. Woo, and H. Yeo. Data-Driven Imputation Method for Traffic Data in Sectional Units of Road Links. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 17, No. 6, 2016, pp. 1762–1771. <https://doi.org/10.1109/TITS.2016.2530312>.
15. Zhong, M., S. Sharma, and Z. Liu. Assessing Robustness of Imputation Models Based on Data from Different Jurisdictions Examples of Alberta and Saskatchewan , Canada. *Transportation research record*, Vol. 1917, No. 1, 2005, pp. 116–126.
16. Ni, D., and J. D. Leonard. Markov Chain Monte Carlo Multiple Imputation Using Bayesian Networks for Incomplete Intelligent Transportation Systems Data. *Transportation Research Record*, No. 1935, 2005, pp. 57–67. <https://doi.org/10.3141/1935-07>.
17. Tipping, M. E., and C. M. Bishop. Mixtures of Probabilistic Principal Component Analyzers. *Neural computation*, Vol. 11, No. 2, 1999, pp. 443–482.
18. Qu, L., L. Li, Y. Zhang, and J. Hu. PPCA-Based Missing Data Imputation for Traffic Flow Volume : A Systematical Approach. Vol. 10, No. 3, 2009, pp. 512–522.
19. Zhong, M., S. Sharma, and P. Lingras. Genetically Designed Models for Accurate Imputation of Missing Traffic Counts. No. 1879, 2004, pp. 71–79.
20. Zhang, C., S. Sun, and G. Yu. A Bayesian Network Approach to Time Series Forecasting of Short-Term Traffic Flows. *The 7th International IEEE Conference on Intelligent Transportation Systems*, No. 04TH8749, 2004, pp. 216–221.
21. Ghosh, B., B. Basu, and M. O’Mahony. Bayesian Time-Series Model for Short-Term Traffic Flow Forecasting. *Journal of transportation engineering*, Vol. 133, No. 3, 2007, pp. 180–189.
22. Dia, H. An Object-Oriented Neural Network Approach to Short-Term Traffic Forecasting. *European Journal of Operational Research*, Vol. 131, No. 2, 2001, pp. 253–261.
23. Vlahogianni, E. I., M. G. Karlaftis, and J. C. Golias. Optimized and Meta-Optimized Neural Networks for Short-Term Traffic Flow Prediction : A Genetic Approach. *Transportation Research Part C: Emerging Technologies*, Vol. 13, No. 3, 2005, pp. 211–234. <https://doi.org/10.1016/j.trc.2005.04.007>.
24. Castro-Neto, M., Y. S. Jeong, M. K. Jeong, and L. D. Han. Online-SVR for Short-Term Traffic Flow Prediction under Typical and Atypical Traffic Conditions. *Expert Systems*

- 1 with *Applications*, Vol. 36, No. 3, 2009, pp. 6164–6173.
2 <https://doi.org/10.1016/j.eswa.2008.07.069>.
- 3 25. Weiss, K., T. M. Khoshgoftaar, and D. Wang. A Survey of Transfer Learning. *Journal of*
4 *Big Data*, Vol. 3, No. 1, 2016, pp. 1–40. <https://doi.org/10.1186/s40537-016-0043-6>.
- 5 26. Yu, J., H. Wei, H. Guo, and Y. Cai. Urban Traffic State Prediction Based on SA-LSTM.
6 *IOP Conference Series: Earth and Environmental Science*, Vol. 783, No. 1, 2021.
7 <https://doi.org/10.1088/1755-1315/783/1/012153>.
- 8 27. Van Der Voort, M., M. Dougherty, and S. Watson. Combining Kohonen Maps with
9 ARIMA Time Series Models to Forecast Traffic Flow. *Transportation Research Part C:*
10 *Emerging Technologies*, Vol. 4, No. 5, 1996, pp. 307–318. [https://doi.org/10.1016/S0968-](https://doi.org/10.1016/S0968-090X(97)82903-8)
11 090X(97)82903-8.
- 12 28. Williams, B. M., and L. A. Hoel. Modeling and Forecasting Vehicular Traffic Flow as a
13 Seasonal ARIMA Process: Theoretical Basis and Empirical Results. *Journal of*
14 *Transportation Engineering*, Vol. 129, No. 6, 2003, pp. 664–672.
15 [https://doi.org/10.1061/\(ASCE\)0733-947X\(2003\)129:6\(664\)](https://doi.org/10.1061/(ASCE)0733-947X(2003)129:6(664)).
- 16 29. Shi, G., J. Guo, W. Huang, and B. M. Williams. Modeling Seasonal Heteroscedasticity in
17 Vehicular Traffic Condition Series Using a Seasonal Adjustment Approach. *Journal of*
18 *Transportation Engineering*, Vol. 140, No. 5, 2014, pp. 1–11.
19 [https://doi.org/10.1061/\(ASCE\)TE.1943-5436.0000656](https://doi.org/10.1061/(ASCE)TE.1943-5436.0000656).
- 20 30. Ojeda, L. L., A. Y. Kibangou, and C. C. De Wit. Adaptive Kalman Filtering for Multi-
21 Step Ahead Traffic Flow Prediction. *Proceedings of the American Control Conference*,
22 2013, pp. 4724–4729. <https://doi.org/10.1109/acc.2013.6580568>.
- 23 31. Guo, J., W. Huang, and B. M. Williams. Adaptive Kalman Filter Approach for Stochastic
24 Short-Term Traffic Flow Rate Prediction and Uncertainty Quantification. *Transportation*
25 *Research Part C: Emerging Technologies*, Vol. 43, 2014, pp. 50–64.
26 <https://doi.org/10.1016/j.trc.2014.02.006>.
- 27 32. Dell’acqua, P., F. Bellotti, R. Berta, and A. De Gloria. Time-Aware Multivariate Nearest
28 Neighbor Regression Methods for Traffic Flow Prediction. *IEEE Transactions on*
29 *Intelligent Transportation Systems*, Vol. 16, No. 6, 2015, pp. 3393–3402.
30 <https://doi.org/10.1109/TITS.2015.2453116>.
- 31 33. Cai, P., Y. Wang, G. Lu, P. Chen, C. Ding, and J. Sun. A Spatiotemporal Correlative K-
32 Nearest Neighbor Model for Short-Term Traffic Multistep Forecasting. *Transportation*
33 *Research Part C: Emerging Technologies*, Vol. 62, 2016, pp. 21–34.
34 <https://doi.org/10.1016/j.trc.2015.11.002>.
- 35 34. Wu, S., Z. Yang, X. Zhu, and B. Yu. Improved K-Nn for Short-Term Traffic Forecasting
36 Using Temporal and Spatial Information. *Journal of Transportation Engineering*, Vol.
37 140, No. 7, 2014, pp. 1–9. [https://doi.org/10.1061/\(ASCE\)TE.1943-5436.0000672](https://doi.org/10.1061/(ASCE)TE.1943-5436.0000672).
- 38 35. Sun, B., W. Cheng, P. Goswami, and G. Bai. Short-Term Traffic Forecasting Using Self-

- Adjusting k-Nearest Neighbours. *IET Intelligent Transport Systems*, Vol. 12, No. 1, 2018, pp. 41–48. <https://doi.org/10.1049/iet-its.2016.0263>.
36. Wang, J., W. Deng, and Y. Guo. New Bayesian Combination Method for Short-Term Traffic Flow Forecasting. *Transportation Research Part C: Emerging Technologies*, Vol. 43, 2014, pp. 79–94. <https://doi.org/10.1016/j.trc.2014.02.005>.
37. Xu, Y., Q. J. Kong, R. Klette, and Y. Liu. Accurate and Interpretable Bayesian MARS for Traffic Flow Prediction. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 15, No. 6, 2014, pp. 2457–2469. <https://doi.org/10.1109/TITS.2014.2315794>.
38. Smith, B., and M. J. Demetsky. SHORT-TERM TRAFFIC FLOW PREDICTION MODELS - A COMPARISON OF NEURAL NETWORK AND NONPARAMETRIC REGRESSION APPROACHES. *Sensors (Peterborough, NH)*, 1994, pp. 1–4.
39. Chen, D. Research on Traffic Flow Prediction Method Based on BP Neural Network. *EEE Transactions on Industrial Informatics*, Vol. 13, No. 4, 2017, pp. 2000–2008.
40. Qu, Z., H. Li, Z. Li, and T. Zhong. Short-Term Traffic Flow Forecasting Method with M-B-LSTM Hybrid Network. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 23, No. 1, 2022, pp. 225–235. <https://doi.org/10.1109/TITS.2020.3009725>.
41. Park, D., L. R. Rilett, and G. Han. Spectral Basis Networks for Real-Time Travel Time Forecasting. *Journal of Transportation Engineering*, Vol. 3, No. December, 1999, pp. 515–523.
42. Van Lint, J. W. C., S. P. Hooqendoorn, and H. J. Van Zuvlen. Freeway Travel Time Prediction with State-Space Neural Networks Modeling State-Space Dynamics with Recurrent Neural Networks. *Transportation Research Record*, No. 1811, 2002, pp. 30–39. <https://doi.org/10.3141/1811-04>.
43. Lingras, P., S. Sharma, and M. Zhong. Prediction of Recreational Travel Using Genetically Designed Regression and Time-Delay Neural Network Models. *Transportation Research Record*, No. 1805, 2002, pp. 16–24. <https://doi.org/10.3141/1805-03>.
44. Gers, F. A., J. Schmidhuber, and F. Cummins. *Lerning to Forget: Continual Prediction with LSTM*. 1999.
45. Hochreiter, S., and J. Urgan Schmidhuber. Long Shortterm Memory. *Neural Computation*, Vol. 9, No. 8, 1997, p. 17351780.
46. Yang, B., S. Sun, J. Li, X. Lin, and Y. Tian. Traffic Flow Prediction Using LSTM with Feature Enhancement. *Neurocomputing*, Vol. 332, 2019, pp. 320–327. <https://doi.org/10.1016/j.neucom.2018.12.016>.
47. Tian, Y., K. Zhang, J. Li, X. Lin, and B. Yang. Neurocomputing LSTM-Based Traffic Flow Prediction with Missing Data. Vol. 318, 2018, pp. 297–305. <https://doi.org/10.1016/j.neucom.2018.08.067>.

48. Fu, R., Z. Zhang, and L. Li. Using LSTM and GRU Neural Network Methods for Traffic Flow Prediction. In *2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)*. IEEE, 2016, pp. 324–328.
49. Luo, X., D. Li, Y. Yang, and S. Zhang. Spatiotemporal Traffic Flow Prediction with KNN and LSTM. *Journal of Advanced Transportation*, Vol. 2019, 2019. <https://doi.org/10.1155/2019/4145353>.
50. Mackenzie, J., J. F. Roddick, and R. Zito. An Evaluation of HTM and LSTM for Short-Term Arterial Traffic Flow Prediction. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 20, No. 5, 2019, pp. 1847–1857. <https://doi.org/10.1109/TITS.2018.2843349>.
51. Kang, D., Y. Lv, and Y. Chen. Short-Term Traffic Flow Prediction with LSTM Recurrent Neural Network. *IEEE 20th International Conference on Intelligent Transportation Systems*, 2017, pp. 1–6.
52. Do, L. N. N., H. L. Vu, B. Q. Vo, Z. Liu, and D. Phung. An Effective Spatial-Temporal Attention Based Neural Network for Traffic Flow Prediction. *Transportation Research Part C: Emerging Technologies*, Vol. 108, No. August, 2019, pp. 12–28. <https://doi.org/10.1016/j.trc.2019.09.008>.
53. Pan, S. J., and Q. Yang. A Survey on Transfer Learning. *IEEE Transactions on knowledge and data engineering*, Vol. 22, No. 10, 2009, pp. 1345–1359.
54. Folgado, D., M. Barandas, R. Matias, R. Martins, M. Carvalho, and H. Gamboa. Time Alignment Measurement for Time Series. *Pattern Recognition*, Vol. 81, 2018, pp. 268–279. <https://doi.org/10.1016/j.patcog.2018.04.003>.
55. Fu, T. C. A Review on Time Series Data Mining. *Engineering Applications of Artificial Intelligence*, Vol. 24, No. 1, 2011, pp. 164–181. <https://doi.org/10.1016/j.engappai.2010.09.007>.
56. Li, H., J. Liu, Z. Yang, R. W. Liu, K. Wu, and Y. Wan. Adaptively Constrained Dynamic Time Warping for Time Series Classification and Clustering. *Information Sciences*, Vol. 534, 2020, pp. 97–116. <https://doi.org/10.1016/j.ins.2020.04.009>.
57. Wu, Y., H. Tan, L. Qin, B. Ran, and Z. Jiang. A Hybrid Deep Learning Based Traffic Flow Prediction Method and Its Understanding. Vol. 90, No. April 2017, 2018, pp. 166–180. <https://doi.org/10.1016/j.trc.2018.03.001>.
58. Chen, M., X. Yu, and Y. Liu. PCNN: Deep Convolutional Networks for Short-Term Traffic Congestion Prediction. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 19, No. 11, 2018, pp. 3550–3559. <https://doi.org/10.1109/TITS.2018.2835523>.