# Long Short-term Memory Neural Network for Short-term High-speed Rail Passenger Flow Forecasting 

Yangyang Zhao ${ }^{\text {a }}$, Xinguo Jiang ${ }^{\text {a, }, 1}$<br>${ }^{\text {a }}$ Department of transportation and logistics, Southwest Jiaotong University<br>National United Engineering Laboratory of Integrated and Intelligent Transportation, West<br>Park, High-Tech District, Chengdu, China 611756<br>${ }^{1}$ E-mail: ejiang@gmail.com, Phone: +86 15229308543


#### Abstract

: The uncertainty of estimating the railway passenger flow in advance may disrupt the passenger operation and management (e.g., passenger evacuation planning, seat allocation, and train timetable programming). In order to proactively improve the service quality and efficiency of the railway system, the short-term passenger flow prediction technique is vital in the field of operation and management system. Utilizing the deep learning library-Keras, the study develops a long short-term memory neural network (LSTM NN) to predict the short-term high-speed rail (HSR) passenger flow. Processing the raw data, we first construct the passenger flow sequences as the input (output) variables. Then the gird search and cross validation techniques are applied to optimize the LSTM NN parameters. At last we utilize the data provided by Shanghai railway administration of China as the case study. Through a comparison with other representative methods, including Auto-Regressive Integrated Moving Average (ARIMA), Back Propagation Neural Network (BPNN), and Support Vector Machine Regression (SVR), results suggest that the proposed LSTM NN can generate great potentials for accurate passenger flow predictions.


## Keywords:

Short-term passenger flow prediction, High-speed rail, Long short-term memory neural network, Grid search, Cross validation

## 1 Introduction

High-speed rail (HSR), as a high-quality inter-city transportation mode, is developing rapidly in many countries. For example, in China HSR has become more and more popular with travelers and can effectively relieve the pressure of transporting passengers among the major metropolises. For railway operators, short-term forecasting is closely related to revenue management and service quality. Demand information provided by short-term forecasting can be used as inputs for other systems such as passenger evacuation, seat location, pricing, and train timetable programming. Thus, accurate prediction of short-term passenger flow is significant in the railway operational decision-making and dynamic operation adjustment. In the past decades, numerous short-term traffic flow prediction models have been proposed in the
transportation systems (e.g., freeway, railway, bus, and metro). These models can be generally categorized into parametric and nonparametric ones.

Parametric models, in particular smoothing techniques (Williams, 1998), grey forecasting model (Hsu and Wen, 1998; Fang and Wu, 2006), state space model (Anthony and Karlaftis, 2003), and autoregressive integrated moving average (ARIMA) (Hamed et al., 1995; Lee and Fambro, 1999) have been used extensively. Especially, ARIMA has aroused wide interest since 1970s due to its effectiveness in modeling linear and stationary time series. Utilizing the 20 -sec ( $30-\mathrm{sec}$ and $60-\mathrm{sec}$ ) traffic flow data, Ahmed and Cook (1979) showed that ARIMA outperformed moving-average, double-exponential smoothing, and exponential smoothing with adaptive response. Lee and Fambro (1999) revealed that the subset ARIMA provided more stable and accurate predictions than full ARIMA through 5-min traffic flow prediction. What's more, with the $15-\mathrm{min}$ traffic flow data, a comparison conducted by Williams and Hoel (2003) between the nearest-neighbor (neural network and historical average model) and seasonal ARIMA favored the seasonal ARIMA. However, ARIMA assumed a linear correlation among time series data and might not address the nonlinearity issue inherent in the traffic flow; comparatively, nonparametric techniques could deal with the nonlinearity and were expected to achieve more accurate predictions. Generally, ARIMA is compared as a benchmarking method to the newly proposed nonparametric models.

For the nonparametric models, much more work has been done such as Bayesian network (Zheng et al., 2006), Kalman filtering (Qkutani and Stephanedes, 1984), Support Vector Machine Regression (SVR) (Manoel et al., 2009), K-nearest Neighbor Model (Zhang et al., 2013), the probability tree (Leng et al., 2013), and the random forest (Kecman and Goverde, 2015). Regarding the short-term prediction of passenger flow only, Wei and Chen (2012) combined empirical mode decomposition (EMD) with back-propagation neural network (BPNN) to predict the $15-\mathrm{min}$ metro passenger flow. Sun et al. (2015) integrated Wavelet with SVM to forecast Beijing subway ridership particularly in the rush hours. Under special events scenarios, Li et al. (2017) developed multiscale radial basis function networks to predict the 15min metro passenger flow. Additionally, gradient boosting decision trees (Ding et al., 2016) was also applied to predict $15-\mathrm{min}$ subway ridership and identify the relative influences of the independent predictor input variables. In addition, to predict railway passenger flow in a day, Tsai et al. (2009) constructed a multiple temporal unit neural network and a parallel ensemble neural network, and Jiang et al. (2014) devised a hybrid model which integrated ensemble EMD with grey support vector machine (GSVM).

Among the nonparametric models, neural networks have drawn the greatest attentions for its mapping capabilities. As a subset of neural network, recently deep learning has been applied with success in many fields, such as dimensionality reduction (Hinton and Salakhutdinov, 2006), natural language processing (Collobert and Weston, 2008), object detection (Goodfellow et al., 2013), and classification tasks (LeCun et al., 2015). Therefore, it inspires us to combine the short-term prediction with the deep architecture models. However, currently most researchers concentrated on road traffic (Liu and Chen, 2017; Bai et al., 2017; Polson and Sokolov, 2017; Mackenzie et al., 2018) and limited attention has been paid to railway passenger flow. The paper develops a LSTM NN to predict the short-term HSR passenger flow and the effectiveness of the proposed LSTM NN is validated through a comparison with ARIMA, BP NN, and SVR.

## 2 Methodology

### 2.1 Long short-term memory neural network

LSTM NN was originally introduced by Hochreiter and Schmidhuber (1970) and improved by Gers at al. (2000). The primary objective of LSTM NN is to model long-term dependencies. A LSTM NN is composed of one input layer, one recurrent hidden layer and one output layer. Different from the traditional NN, the basic unit of the hidden layer is memory block (Abigogun, 2005). The memory block contains a memory cell, an input gate, an output gate, and a forget gate. The cell is responsible for transporting values over arbitrary time intervals and memorizing the temporal state. Three gates can be treated as "conventional" artificial neurons, similar to those in a feedforward neural network (i.e., the input gate and output gate control the input and output activations into the block, the forget gate selects the partial output from the upper memory block to prevent the cell values growing without bound). Through the multiplicative gates, LSTM memory cells can store and access information during the long periods of time, thus mitigating the vanishing gradient problem. The above procedure is shown in Figure 1.


Figure 1: LSTM NN architecture
Given the model input $x=\left(x_{1}, x_{2}, \cdots, x_{T}\right)$, the output $y=\left(y_{1}, y_{2}, \cdots, y_{T}\right)$, and the hidden output $h=\left(h_{1}, h_{2}, \cdots, h_{T}\right)$, where $T$ represents the prediction period. In the context of shortterm passenger flow prediction, $x$ can be considered as historical input data (e.g., time of day, weather condition, passenger flow), and $y$ is the estimated passenger flow. The predicted passenger flow will be iteratively calculated by equations (1)-(8):

$$
\begin{align*}
& i_{t}=\sigma\left(W_{x i} x_{t}+W_{h i} h_{t-1}+b_{i}\right)  \tag{1}\\
& f_{t}=\sigma\left(W_{x f} x_{t}+W_{h f} h_{t-1}+b_{f}\right)  \tag{2}\\
& c_{t}=f_{t} * c_{t-1}+i_{t} * \tanh \left(W_{x c} x_{t}+W_{h c} h_{t-1}+b_{c}\right)  \tag{3}\\
& o_{t}=\sigma\left(W_{x o} x_{t}+W_{h o} h_{t-1}+b_{o}\right)  \tag{4}\\
& h_{t}=o_{t} * \tanh \left(c_{t}\right)  \tag{5}\\
& y_{t}=\sigma\left(W_{h y} h_{t}+b_{y}\right) \tag{6}
\end{align*}
$$

where $t$ is the order of observation time interval during $T, W$ terms are weight matrices (e.g., $W_{x i}$ is the matrix of weights from the input gate $i$ to the input $x$ ), the $b$ terms are bias vectors
( $b_{i}$ is the input gate $i$ bias vector), and $i, f, o, c$ represent the input gate, forget gate, output gate, and cell activation vectors, respectively. $\delta(\cdot)$ is the standard logistic sigmoid function:

$$
\begin{equation*}
\sigma(x)=\frac{1}{1+\mathrm{e}^{-x}} \tag{7}
\end{equation*}
$$

The square error is used as the loss function as follows:

$$
\begin{equation*}
e(t)=\sum\left(y_{t}-\bar{y}_{t}\right)^{2} \tag{8}
\end{equation*}
$$

where $y_{t}, \bar{y}_{t}$ is the observed and predicted passenger flow, respectively.
The truncated Back Propagation Through Time (BPTT) is widely used to train LSTM NN. Due to extensive mathematical derivations, the detailed steps are not covered in this section and readers may refer to (Gers, 2001) for more information.

### 2.2 Other representative models

We employ three typical forecasting models: ARIMA, BPNN, and SVR to test the performance of the LSTM NN.

## ARIMA

ARIMA models are linear estimators regress on past values of the modeled time series (the autoregressive terms) or past prediction errors (the moving average terms), and are also written as ARIMA $(p, d, q)$, where $p$ is the number of autoregressive terms, $d$ is the number of order, and $q$ is the moving average parameter.

## BPNN

BPNN is a kind of artificial neural network which adopts a backpropagation algorithm to modify the weights of the neurons, and thus minimize the errors between the actual output values and the target output values. The basic structure of BPNN consists of an input layer, a hidden layer, and an output layer. Details of the algorithm is described in (Park and Rilett, 1999).

## SVR

SVR is SVM for regression, it is based on the computation of a linear regression function in a high dimensional feature space, where the input data are mapped via a non-linear function. Several studies have shown the SVR effectiveness in forecasting traffic flow (Chen et al., 2011; Zhang and Liu, 2009; Zhang and Xie, 2007).

### 2.3 Grid search and cross validation

Grid search is the process of scanning the data to configure the optimal parameters for a given model. Considering the possible values of models, grid search will build a model on each parameter combination and iterate through each combination accordingly. With different model performance, it can eventually select the optimal parameters of the given model. It needs to note that grid search can be extremely computationally expensive when dealing with a high dimensional set of parameters. Generally, parameters in the LSTM NN consist of the training batch-size, the epoch, the activation function of each hidden layer, the number of hidden layers and the hidden neurons.

Cross validation is a class of model evaluation methods. The basic idea is that some of the data is removed from the entire data set before training, and the removed data can be used to test the performance of the learned model on "new" data. According to the different training set and testing sets, there are mainly three kinds of cross validation, including the holdout method, $K$-fold cross validation, and leave-one-out cross validation. In the paper $K$-fold cross validation is adopted and the process is listed in as follows:

Step1. Dividing the data set into $k$ subsets.
Step2. Taking one of the $k$ subsets as the test set and the other $k-1$ subsets as the training set, the test error is recorded.

Step3. Training and testing the model for $k$ times to ensure every data point gets to be in the test sets exactly once and in the training sets $k-1$ times.

Step4. Evaluating the model by calculating the average test error.
Grid search and cross validation are usually integrated together to find the model's optimal parameters and evaluate its performance on "new" data simultaneously (Anguita et al., 2009; Krstajic et al., 2014).

## 3 Experimentation

The short-term passenger flow prediction problem can be stated as follows. Let $\left\{x_{t}^{T}\right\}$ denote the observed passenger flow in the $t^{\text {th }}$ time interval of the $T^{\mathrm{h}}$ day during a period time. Given a sequence $\left\{x_{t}^{T}\right\}$ of observed passenger flow, $T=T, T-1, \ldots, T-\mathrm{m}$, the goal is to predict the $\left\{x_{t}^{T+N}\right\}$, where $N$ represents the prediction horizon and $m$ represents the length of observation time period.

### 3.1 Data description

The daily sale data of Shanghai-Beijing HSR line from July 1 (Saturday), 2017 to July 31 (Monday), 2017 are provided by the Shanghai railway administration of China. There are 24 stations along Beijing-Shanghai HSR line with the length of $1,318 \mathrm{~km}$. Table 1 lists the related data fields.

Table 1: List of data fields

| No | Field | Description |
| :--- | :--- | :--- |
| 1 | Transcation_id | Identifying a transcation |
| 2 | Transcation_timestamp | Time of day, day of week |
| 3 | Station_name | Origin station, destination station |

Considering the dispatched and attracted HSR passenger flows at each station, three OD pairs with the highest passenger demand include Shanghai-Beijing, Shanghai-Nanjing, and Nanjing-Beijing during the survey period. Taking July 1 as an example, these OD pairs consist of $23.42 \%$ of the total Shanghai-Beijing HSR travel demand $(10.04 \%, 8.27 \%$, and $5.11 \%$, respectively).

### 3.2 Model development

## Training data

The training data consists of input data and output data. In addition to passenger flow $X_{t}^{T}=$ $\left(x_{t}^{T}, x_{t}^{T-1}, \cdots, x_{t}^{T-m}\right)$, the features of time of day $t$, day of week $T$ are all used as inputs in the model training stage, and the output is $x_{t}^{T+N}$. The validation dataset has the same features as the training dataset, and the test dataset has the same features as the inputs of training dataset, the output is the target value for prediction. Table 2 describes the code values of each feature.

Table 2: Feature descriptions for training data and label data

| No | Feature | Code values | Description |
| :---: | :--- | :---: | :---: |
| 1 | Time of day $(t)$ | $0-11$ | $0-11$ represent the 12 hours in |
| 2 | Day of week $(T)$ | $1-7$ | $6: 00-18: 00$ |
| 3 | Passenger flow | $x_{t}^{T}$ | $1-7$ represent Monday-Sunday |
| hourly aggregated passenger flow |  |  |  |

Based on the collected daily sale data, normally we have to normalize the training data by equation (9) to improve the model efficiency:

$$
\begin{equation*}
\bar{x}=(x-a) /(c-b) \tag{9}
\end{equation*}
$$

where $x$ denotes the code value of a feature, $\mathrm{a}, \mathrm{b}, \mathrm{c}$ is the average, minimum and maximum value of a feature, respectively, the normalized $X_{t}^{T}, x_{t}^{T}, T, t$ is marked as $\left(X_{t^{\prime}}^{T^{\prime}}, x_{t^{\prime}}^{T^{\prime}}, T^{\prime}, t^{\prime}\right)$. As a result, for the LSTM NN, the final input is a sequence of passenger flow $x_{t^{\prime}}^{T^{\prime}}$ with its corresponding temporal features $T^{\prime}$ and $t^{\prime}$, that is, vectors $X_{t^{\prime}}^{T^{\prime}}, \boldsymbol{T}^{\prime}$ and $\boldsymbol{t}^{\prime}$, respectively, and the output is $x_{t^{\prime}}^{T^{\prime}+1}$.

## Model structure

The first 28-day (July 1-28) data are used for training (90\%) and validation (10\%), and the last 3-day (July 29-31, Saturday-Monday) data for testing. The prediction time ranges from 06:0018:00, time interval is one hour. One-step ahead prediction means prediction horizon $N$ is 1 . HSR passenger flow shows regular changes in weeks, thus the length of observation time period $m$ is 7. Given the small datasets in the paper, we fuse the grid search and the $k$-fold cross validation to enhance the model stability and reliability. The $k$ is $3 \sim 10$ empirically due to massive calculation. Here we set $k$ as 7 , which means to train the model 7 times on different training and validation data sets. According to related studies(Liu and Chen, 2017; Lv et al., 2015), the hidden layer (LSTM layer) size ranges from 1 to 6 , the number of hidden units $l_{N} \in\{30,60,90,120,150,180\}$, the activation function of each layer is "tanh", the batch-size $l_{S} \in\{1,2,3,4\}$, epoch value depends on the number of training parameters based on equations (1-7), at last a dense layer is added to the output layer. Figure 2 shows the prediction process.

The deep learning library-Keras is a high-level neural networks API, written in Python and capable of running on top of TensorFlow, CNTK, or Theano. It was developed with a focus on enabling fast experimentation. Utilizing the Keras, Table 3 shows the optimal parameters in the LSTM NN.


Figure 2: The training and forecasting process
Table 3: The key hyperparameters in the LSTM NN

| Task | Hidden layers | Hidden units(bottom-top) |
| :---: | :---: | :---: |
| A day ahead passenger flow prediction <br> (time interval : an hour) | 2 | $[90,90]$ |

For the ARIMA, the input is passenger flow vector $X_{t}^{T}$ and the output is $x_{t}^{T+1}$, the best model ARIMA $(5,0,2)$ selected by auto.arima function.

For the BPNN a, the input is $X_{t^{\prime}}^{T^{\prime}}, \boldsymbol{T}^{\prime}, \boldsymbol{t}^{\prime}$ and the output is $x_{t^{\prime}}^{T^{\prime}+1}$, the grid search and the cross-validation were utilized to build the optimal structure of the BP NN, while the hidden layer size was less than 3 caused by vanishing gradient.

For the SVR, the input is $X_{t^{\prime}}^{T^{\prime}}, \boldsymbol{T}^{\prime}, \boldsymbol{t}^{\prime}$ and the output is $x_{t^{\prime}}^{T^{\prime}+1}$, radial basis function (RBF) was used (Cherkassky and Ma, 2004) with three other parameters: cost $C$, width parameter $g$, and epsilon $\varepsilon$. Parameters were learned by parameter tuning function tune using gird search.

### 3.3 Model evaluation

The prediction accuracy is evaluated with the use of Mean Absolute Percentage Error (MAPE) and Root Mean Square error (RMSE). These measures are defined as follows:

$$
\begin{align*}
& \operatorname{MAPE}(y, \hat{y})=\frac{1}{n} \sum_{i=1}^{n} \frac{\left|y_{i}-\hat{y}_{i}\right|}{y_{i}}  \tag{10}\\
& \operatorname{RMSE}(y, \hat{y})=\left[\frac{1}{n} \sum_{i=1}^{n}\left(y_{i}-\hat{y}_{i}\right)^{2}\right]^{\frac{1}{2}} \tag{11}
\end{align*}
$$

where $y_{i}$ is the observed passenger flow, $\hat{y}_{i}$ is the predicted passenger flow, and $n$ is the total number of predictions.

## 4 Results

The OD pairs (Shanghai-Beijing and Shanghai-Nanjing) with the highest passenger demand belong to two typical categories (i.e., long distance (more than 800 km ) and middle distance $(200-800 \mathrm{~km}))$ in the railway operations, which are selected for the prediction. Figure 3 shows the outputs of ARIMA, BPNN, SVR, and LSTM NN. Table 4 presents the forecasting errors.



Figure 3: Prediction results from July 29 to 31 (a: Shanghai-Beijing; b: Shanghai-Nanjing)
Table 4: Prediction accuracy comparison

| OD | Model | MAPE | RMSE | OD | Model | MAPE | RMSE |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Shanghai- | LSTM | $8.01 \%$ | 133 |  | LSTM | $8.47 \%$ | 73 |
|  | SVR | $11.79 \%$ | 187 | Shanghai- | BP | $11.96 \%$ | 91 |
|  | ARIMA | $15.17 \%$ | 198 | Nanjing | SVR | $14.39 \%$ | 117 |
|  | $16.49 \%$ | 204 |  | ARIMA | $16.25 \%$ | 123 |  |

According to the MAPE and RMSE values in Table 4, the LSTM NN is superior to the BPNN, SVR, and ARIMA for the short-term HSR passenger flow prediction, while the BPNN outperforms SVR, and SVR outperforms ARIMA. The results strongly indicate that the relationship between the historical and forecasting passenger flow are non-linearly correlated.

Also, Figure3 shows that passenger flow time distribution is significantly different in the two chosen OD pairs. Additionally, the test data include passenger flow of both working and non-working days (Saturday-Monday) between which passenger flow time distribution also shows a weak regularity for both OD pairs, however, the prediction accuracy of four models shows slight fluctuations. The observations can be attributed to two possible reasons. On the one hand, the adopted models are efficient enough to forecast passenger flow. On the other hand, given the chosen OD pairs, the mathematical relationship between the input and output is similar, which leads to the stable model performance.

Nonetheless, the LSTM NN does not seem to perform well in several time intervals (e.g., 11:00-12:00 on July 29 and 15:00-16:00 on July 30 (Shanghai-Beijing);16:00-17:00 on July 30 and July 31 (Shanghai-Nanjing)) in Figure 4.

(a)

 ódo
 July (29-31)
Figure 4: Prediction errors analysis (Shanghai-Beijing, Shanghai-Nanjing)

The phenomenon may be explained that the model input features (i.e., time of day, day of week, and passenger flow volume) are not sufficient for prediction under special scenarios. For instance, passenger flow with a sharp increase (e.g., 13:00-14:00 on July 30(Shanghai-Beijing)) or decrease (e.g., 16:00-17:00 on July 31(Shanghai-Nanjing)) was easily affected by other factors (e.g., weather and an emergency), we have not considered these factors yet and consequently degraded the model prediction performance. More features need to be assessed and added to enhance the model reliability and stability. Besides, to analyze the LSTM NN performance with different time intervals, $x_{t}^{T}$ is also aggregated by 2-hour (3-hour, 4-hour, and 5-hour), Figure 5 shows the forecasting errors.


Figure 5: Prediction errors with different time intervals
With larger prediction time interval, the forecasting errors of both OD pairs become smaller in Figure 5. The underlying reason is that smaller time interval increases the fluctuation of passenger flow data, while greater interval contributes to more stability and thus passenger flow prediction is relatively easier. The conclusion is consistent with the variability in regularity of metro passenger flow (Zhong et al., 2016), it points out that dramatically increased invariability may occur up to the temporal scale of about 15 minutes according to the cases of Beijing, Singapore, and London, implying that time interval limits exist when we attempt to forecast the short-term passenger flow.

Additionally, we evaluate the forecasting performance in up to 7 -step ahead, that is, to predict $x_{t}^{T+1}, x_{t}^{T+2}, x_{t}^{T+3}, x_{t}^{T+4}, x_{t}^{T+5}, x_{t}^{T+6}, x_{t}^{T+7}$, respectively. We first forecast the first step, then the forecasted value is applied as a part of the input variables for the next step prediction. Thus, the entire horizon can be repeatedly predicted step by step. Figure 6 shows the forecasting errors.


Figure 6: Prediction errors with different forecasting steps
As the forecasting step increases, the prediction accuracy also decreases in Figure 6. When $N>3$, the performance of the LSTM NN degrades significantly. The reason could be that the cumulative forecasting error causes the input data to be much less efficient, and $x_{t}^{T+N}$ is less predictable.

## 5 Conclusions

The main contribution of the paper is to develop a LSTM model to predict short-term HSR passenger flow. Preprocessing the raw data, we construct the input and output variables, and get the optimal LSTM parameters through grid search and cross-validation. Compared with the traditional forecasting methods (ARIMA, BPNN, and SVR), the proposed model poses a great potential for short-term passenger flow prediction. In addition, the LSTM NN performed better with larger time intervals, but performance degrades significantly with the increase of forecasting step.

One of the drawbacks of deep learning models is the low explanatory power (Polson and Sokolov, 2017)). In a recent review of short-term forecasting technique (Vlahogianni et al., 2014), model interpretability is mentioned as one of the barriers in implementing more sophisticated machine learning models in practice. Liu and Chen (2017) tried to explain that the deep architecture could extract the deep features and perform well in BRT passenger flow prediction, while the passenger travel behavior of BRT is different from HSR. The issue needs to be studied further.

## Acknowledgements

Shanghai railway administration of China is acknowledged for providing the data, and they own the data used in the present paper. The authors would like to acknowledge the support from National Natural Science Foundation of China (No.71771191).

## References

Abigogun, O.A., 2005. Data mining, fraud detection and mobile telecommunications: Call pattern analysis with unsupervised neural networks, Master Thesis, University of the Western Cape.
Ahmed, M.S., and Cook, A.R., 1979. "Analysis of freeway traffic time-series data by using BoxJenkins techniques", Transportation Research Record: Journal of the Transportation Research Board, vol. 773, pp. 1-9.
Anguita, D., Ghio, A., Ridella,S., Sterpi, D., 2009. "K-fold cross validation for error rate estimate in support vector machines", In: Proceedings of The International Conference on Data Mining, Austin, United States.
Anthony, S., Karlaftis, M.G., 2003. "A multivariate state space approach for urban traffic flow modeling and prediction", Transportation Research Part C: Emerging Technologies, vol. 11, pp. 121-135.
Bai, Y., Sun, Z., Zeng, B., Deng, J., Li, C, 2017. "A multi-pattern deep fusion model for shortterm bus passenger flow forecasting", Applied Soft Computing, vol. 58, pp. 669-680.
Chen, Q., Li, W., Zhao, J., 2011. "The use of LS-SVM for short-term passenger flow prediction", Transport, vol. 26, pp. 5-10.
Cherkassky,V., Ma, Y. 2004. "Practical selection of svm parameters and noise estimation for svm regression", Neural networks, vol. 17, pp. 113-126.
Collobert, R., and J. Weston., 2008. "A unified architecture for natural language processing: Deep neural networks with multi-task learning", In: Proceedings of The 25th International Conference on Machine Learning, Helsinki, Finland.
Ding, C., Wang, D., Ma, X., Li. H., 2016. "Predicting short-term Subway ridership and prioritizing its influential factors using gradient boosting decision trees", Sustainability, vol. 8, pp. 1100.
Fang, L.J., Wu, Z., 2006. "Application of GM $(1,3)$ in highway passenger capacity forecast of transportation system", Journal of Highway \& Transportation Research \& Development, vol. 26, pp. 163-166.

Gers, F., 2001. Long Short-term Memory in Recurrent Neural Networks, Ph.D. Dissertation, École Polytechnique Fédérale De Lausanne.
Gers, F. A., Schmidhuber, J., Cummins, F., 2000. "Learning to forget: Continual prediction with LSTM", Neural Computation, vol. 12, pp. 2451-2471.
Goodfellow, I.J., Bulatov, Y., Ibarz, J., Arnoud, S., Shet, V., 2013. "Multi-digit number recognition from street view imagery using deep convolutional neural networks", ArXiv Preprint arXiv:1312.6082v4.
Hamed, M.M., AI-Masaeid, H.R., Bani Said, Z.M., 1995. "Short-term prediction of traffic volume in urban arterials", Journal of Transportation Engineering, vol. 121, pp. 249-254.
Hinton, G.E., Salakhutdinov, R.R., 2006. "Reducing the dimensionality of data with neural networks", Science, vol.313, pp. 504-507.
Hochreiter, S., Schmidhuber, J., 1997. "Long short-term memory", Neural Computation, vol. 9, pp. 1735-1780.
Hsu, C.I., Wen, Y.H., 1998. "Improved grey prediction models for the trans-pacific air passenger market", Transportation Planning \& Technology, vol 22, pp. 87-107.
Jiang. X., Zhang. L., and Chen. X., 2014. "Short-term forecasting of high-speed rail demand: a hybrid approach combining ensemble empirical mode decomposition and gray support vector machine with real-world applications in China", Transportation Research Part C: Emerging Technologies, vol. 44, pp. 110-127.
Kecman, P., Goverde, R. M. P., 2015. "Predictive modelling of running and dwell times in railway traffic", Public Transport, vol. 7, pp.295-319.
Krstajic, D., Buturovic, L. J., Leahy, D. E., Thomas, S., 2014. "Cross-validation pitfalls when selecting and assessing regression and classification models", Journal of Cheminformatics, vol. 6, pp. 10.
LeCun, Y., Bengio, Y., Hinton, G.E., 2015. "Deep learning", Nature, vol. 521, pp. 436-444.
Lee, S., Fambro, D.B., 1999. "Application of subset autoregressive integrated moving average model for short-term freeway traffic volume forecasting", Transportation Research Record: Journal of the Transportation Research Board, vol. 1678, pp. 179-188.
Leng, B, Zheng, J.B, Xiong, Z., Lv. W.F, Wan. Y.L., 2013. "Probability tree-based passenger flow prediction and its application to the Beijing subway system", Frontiers of Computer Science, vol. 7, pp. 195-203.
Li, Y., Wang, X., Sun, S., Ma, X., Lu. G., 2017. "Forecasting short-term subway passenger flow under special events scenarios using multiscale radial basis function networks", Transportation Research Part C: Emerging Technologies, vol. 77, pp. 306-328.
Liu, L., Chen, R., 2017. "A novel passenger flow prediction model using deep learning methods", Transportation Research Part C: Emerging Technologies, vol. 84, pp.74-91.
Lv, Y.S., Duan, Y.J., Kang, W.W., Li, Z.X, Wang, F.Y., 2015. "Traffic flow prediction with big data: a deep learning approach", Intelligent transportation systems, vol.16, pp. 865-873.
Mackenzie, J., Roddick, J.F., Zito, R., 2018. "An evaluation of HTM and LSTM for short-term arterial traffic flow prediction", Intelligent transportation systems, pp.1-11.
Manoel, C.N., Jeong, Y.S., Jeong, M.K., Lee, D.H., 2009. "Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions", Expert Systems with Applications, vol. 36, pp. 6164-6173.

Okutani, I., and Stephanedes, Y.J., 1984. "Dynamic prediction of traffic volume through Kalman filtering theory", Transportation Research Part B: Methodological, vol.18, pp. 1-11.
Polson, N.G., Sokolov, V.O., 2017. "Deep learning for short-term traffic flow prediction", Transportation Research Part C: Emerging Technologies, vol. 79, pp. 1-17.
Park, D., Rilett, L.R., 1999. "Forecasting freeway link travel times with a multilayer feedforward neural network", Computer-Aided Civil and Infrastructure Engineering, vol. 14, pp. 357-367.
Sun, Y, Leng, B., Guan, W., 2015. "A novel wavelet-SVM short-time passenger flow prediction in Beijing subway system", Neurocomputing, vol. 166, pp. 109-121.
Tsai, T., Lee, C., Wei, C., 2009. "Neural network based temporal feature models for short-term railway passenger demand forecasting", Expert Systems with Applications, vol. 36, pp. 37283736.

Vlahogianni, E.I., Karlaftis, M.G., Golias, J.C., 2014. "Short-term traffic forecasting: Where we are and where we're going", Transportation Research Part C: Emerging Technologies, vol. 43, pp. 3-19.
Wei, Y., Chen, M.C., 2012. "Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks", Transportation Research Part C: Emerging Technologies, vol. 21, pp. 148-162.
Williams, B.M., Durvasula, P.K., Brown D.E., 1998. "Urban freeway traffic flow prediction: Application of seasonal autoregressive integrated moving average and exponential smoothing models", Transportation Research Record: Journal of the Transportation Research Board, vol. 1644, pp. 132-141.
Williams, B.M., Hoel, L.A., 2003. "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results", Journal of Transportation Engineering, vol. 129, pp. 664-672.
Zhang, L., Liu,Q.C., Yang,W.C., Wei, N., Dong, D.C., 2013. "An improved K-nearest neighbor model for short-term traffic flow prediction", Procedia-Social and Behavioral Sciences, vol. 96, pp. 653-662.
Zhang, Y., Liu, Y., 2009. "Traffic forecasting using least squares support vector machines", Transportmetrica, vol, 5, pp.193-213.
Zhang, Y., Xie, Y., 2007. "Forecasting of short-term freeway volume with v-support vector machines", Transportation Research Record: Journal of the Transportation Research Board, vol. 2024, pp. 92-99.
Zheng, W.Z., Lee, D.H., Shi, Q.X., 2006. "Short-term freeway traffic flow prediction: Bayesian combined neural network approach", Journal of Transportation Engineering, vol. 132, pp. 114-121.
Zhong. C., Batty, M., Manley, E., Wang, Z., Chen, F., 2016. "Variability in regularity: Mining temporal mobility patterns in London, Singapore and Beijing using smart-card data", Plos One, vol. 11, e0149222.

