ORIGINAL PAPER

Deep Learning in Transport Studies: A Meta-analysis on the Prediction Accuracy

Varun Varghese[1,2] · Makoto Chikaraishi[1,2] · Junji Urata[2,3]

Abstract

Deep learning methods are being increasingly applied in transport studies, while the methods require modellers to go through a try-and-error model tuning process particularly on choosing neural network structure. Moreover, the accuracy level also depends on other factors such as the type of data, sample size, region of data collection, and time of prediction. To efficiently facilitate such a model tuning process, this study attempts to summarize the relationship between the prediction accuracy of deep learning models and the factors which influence it. We conducted a comprehensive review of the literature by adopting a detailed search strategy, followed by a meta-analysis on prediction accuracy. Four separate linear mixed effects models, taking into account unobserved heterogeneities in prediction accuracy across studies, were developed to statistically test the impacts of influential factors on prediction accuracy for (a) all observations (136 studies; 2314 cases), (b) studies with MAPE, MRE, and average accuracy indicators (86 studies; 1,878 cases), (c) classification-based studies with accuracy indicator (29 studies; 220 cases), and (d) traffic forecasting studies with MAPE, MRE, and average accuracy indicators (36 studies, 991 cases). The final model includes additional factors to test the influence of sample size and time horizon of prediction variables. The findings showed that, as expected, deep learning models, particularly ones that consider spatiotemporal dependencies of transport phenomena, show better prediction accuracies compared to conventional machine learning models. We also found that, on average, the prediction accuracy is improved by 5.90% with 100 million additional data, while the accuracy is reduced by 5.28% with 100 min increase in time horizon of prediction in traffic forecasting studies. We concluded this paper with a comprehensive summary of the existing findings on the applications of deep learning to transport studies.

Keywords Review paper · Meta-analysis · Deep learning in transportation · Machine learning · Prediction accuracy

Introduction

Challenges in Applying Deep Learning Methods

Artificial intelligence (AI) systems are characterised by their capability to acquire their own knowledge, by understanding patterns from raw data (Goodfellow et al. 2016). The two

✉ Makoto Chikaraishi
chikaraishim@hiroshima-u.ac.jp

[1] Hiroshima University, Higashi-Hiroshima, Hiroshima, Japan

[2] RIKEN Center for Advanced Intelligence Project, Chuo, Tokyo, Japan

[3] The University of Tokyo, Bunkyo, Tokyo, Japan

key components of AI systems are (1) data from sensors, and (2) algorithms/models that translate the data into useful information. If we consider the above-mentioned components, we could say that transportation researchers and practitioners have long been working on AI applications. For example, to control traffic signals, a number of loop detectors have been installed and signals have been automatically controlled following certain models and algorithms in many cities from developed countries (Zhao et al. 2011). Loop detector and global positioning system (GPS) data have also been utilised to provide route guidance information together with the algorithm translating raw data into travel-related information. Thus, AI applications are not new in transportation field. However, the landscape is going to change drastically, mainly because of (1) dramatic increase in data streams from various sensors, and (2) rapid development of machine learning techniques including deep learning.

Conventional machine learning techniques such as random forest method, support vector machines (SVM), and shallow neural network methods have been commonly used in the field of transportation. They have been applied earlier in the fields of traffic state prediction such as traffic speed and flow (Do et al. 2019; Wang and Shi 2013), travel time prediction (Wu et al. 2004), bus arrival time prediction (Bin et al. 2006), transportation mode extraction (Shafique and Hato 2015) among other applications. The last decade has witnessed a rise in the availability of big data in the field of transportation. Data from various sources such as GPS, loop detectors, closed circuit television (CCTV) cameras are being increasingly used in transportation analysis. However, analysing data pertaining to transportation involves the disentanglement of multiple factors of variation. For example, for the prediction of driving behaviour using CCTV or any camera footage, the factors of variation might include several high-level, abstract features such as age and gender of the driver, the various sitting position they can drive in, and the different types of activities they can perform while driving. Achieving high accuracy in such complex situations is a major challenge and deep learning methods are becoming more popular in such scenarios.

Goodfellow et al. (2016) argued that in cases where a nearly human-level understanding of the data is required, representation becomes a major challenge. Deep learning models provide an excellent way to solve this problem by building complex representations based on a combination of simpler representations. Multiple layers can be added to represent complex and abstract features (LeCun et al. 2015), thereby improving the overall accuracy levels. However, with respect deep learning's use in transportation, there remains at least three challenges for researchers and policy-makers. First, the black box functionality of deep learning models is a major barrier in linking these models with the existing transport theories. Policy-makers from across the world have argued on the logic and the reasoning behind using deep learning models (European Parliamentary Research Service 2019; Ministry of Internal Affairs and Communications 2019). Improving accuracy cannot be the lone goal of such models and in principle, these models should be rooted in existing theories. For example, Chikaraishi et al. (2020) argued that, in short-term traffic state prediction, a machine learning model which produces the best prediction accuracy is not always the best for practical use since it does not mimic the mechanisms of congestion occurrence. Second, these methods are cost and resource intensive, with a set of influences from external factors such as type of data and sample size. Traffic forecasting studies, for example, require a huge cost in collecting observed data, as a practical application often needs the collection of city-scale data. Unlike other practical applications of AI, where data could be obtained from a laboratory or a plant,

transportation studies offer a unique challenge with respect to controlling external factors. Third, the decision of choosing appropriate methods often require modellers to adopt a try-and-error approach. The choice of the type of neural network structure would influence the prediction accuracy, which would further depend on the variable being predicted. With multiple areas of applications in transportation, this could get very challenging.

This study intends to contribute to the second and third challenges through a review on the relationship between external factors and the prediction accuracy of deep learning models. A proper summary of existing findings would provide a good guide to reduce burdens in the try-and-error model tuning process, while accounting for other factors such as type of data and sample size.

A Brief Overview of Existing Studies

Currently, both computer scientists and transportation engineering professionals have applied deep learning methods to predict complex transportation phenomena, and their use in transportation studies is rapidly increasing. Initial studies mostly focused on image detection pertaining to the detection of traffic signs (CireşAn et al. 2012), vehicles (Chen et al. 2014), and pedestrians (Ouyang and Wang 2013). However, lately deep learning methods have been applied in many studies analysing complex variables such as traffic state prediction (Bai and Chen 2019; Jo et al. 2019), travel demand estimation (Tang et al. 2019a, b), mode choice and activity choice prediction (Zhao et al. 2019a, b, c, d, e, f).

LeCun et al. (2015) categorised the deep learning methods into three major groups, (a) multilayer architecture using backpropagation (Bengio et al. 2007), (b) convolutional neural networks (CNN) (Simonyan and Zisserman 2014), and c) recurrent neural networks (RNN) (Graves et al. 2013). The use of a type of method depends on the field of application. For example, CNNs have been successful in the analysis of image data and the recognition of its features. Meanwhile, RNN methods have been useful in text and word recognition and in data which required processing of values in sequences (LeCun et al. 2015). Moreover, there exists significant levels of variations within a type of method, for example, the classical RNN structures have evolved into gated RNNs such as the long short-term memory (LSTM) (Hochreiter and Schmidhuber 1997) and the gated recurrent unit (GRU) models (Chung et al. 2014). There are various travel-related variables that could be predicted using deep learning models and the choice of method would significantly influence the prediction accuracy.

Review studies on the application of deep learning in transportation have focused on identifying the areas of application (Nguyen et al. 2018; Wang et al. 2019) and the different kinds of methods (Do et al. 2019). Wang et al. (2019)

identified that deep learning models in transportation have been applied mostly for either classification of discrete states or regression of continuous real values. They identified the areas of application, while describing the advantages and disadvantages of the methods. However, the decision on selecting the type of method and their accuracy would depend upon other external factors such as sample size, area of application, region of data collection (whether from urban area, rural area, or both), and time horizon of prediction (for example, short-term prediction or long-term prediction). Previous review studies did not study accuracy's relationship with respect to the area of application, the type of deep learning method and other external factors such as type and source of data, data coverage, sample size, and time horizon of prediction.

Objectives

To fill in the above-mentioned research gaps, this paper aims (a) to identify the set of deep-learning methods used with respect to the area of application, type of data collected, coverage of the study, time horizon of prediction, and sample size; (b) to statistically determine the relationship of these external factors with prediction accuracy through a meta-analysis. We believe that this review paper will contribute to the existing literature on the application of AI in transportation in two major ways. First, as deep learning is a relatively new field and the number of papers being published have been continuously increasing every year, an extensive survey will cover new studies which were not reviewed in earlier conducted similar works (Do et al. 2019; Nguyen et al. 2018; Wang et al. 2019). Second, it will add on to the understanding of how various factors pertaining to the methodology, data type, coverage area, time of prediction, or sample size are associated with the accuracy. The knowledge about the same will be beneficial for researchers and transportation practitioners who will be able to control for these factors in future studies.

This article is organised as follows. Section 2 describes the research methodology. It has two sub-sections. The first sub-section details out the study selection and search strategy employed to select the papers that were reviewed. Meanwhile, the second sub-section discusses the approach adopted for the meta-analysis. Section 3 discusses the descriptive statistics of the review analysis, which is followed by the discussions of the results of the meta-analysis in Sect. 4. Finally, we conclude our work in Sect. 5 by discussing the key findings, contributions, limitations, and the directions for future research in the field of AI and transportation.

Research Methodology

This section describes the investigation methodology adopted for this study and is further divided into two sub-sections. Section 2.1 illustrates the search strategy implemented to select the papers which used deep learning in transportation analysis. Meanwhile, Sect. 2.2 describes the method and the set of variables considered in the meta-analysis.

Search Strategy

The literature for the review was selected in two steps. First, a generic search on the web of science (WoS) database helped in identifying three review papers already published, which reviewed deep learning studies in the field of transportation. The reference section of these review papers was searched to identify the relevant literature. As a result, a total of 72 unique studies were identified (see Table 1). A conscious decision was made to limit our review to only those studies which analysed travel-related or travel behaviour variables. Therefore, studies which employed deep learning models for vehicle, traffic sign, pedestrian, and cracks in pavement detection were not included in this review. The mentioned areas of application mostly implemented widely available image-based datasets and typically employed different variants of the CNN model (Li et al. 2016; Luo et al. 2014; Qian et al. 2015). The three review studies, viz. Wang et al. (2019), Do et al. (2019) and Nguyen et al. (2018) succinctly summarised the work done so far based on the areas of application, which informed the second step of the search strategy (Wee and Banister 2016). Relevant keywords were identified and systematically searched using the WoS database. The list of keywords is provided in Table 1. Four separate indexes (1) Science Citation Index Expanded (SCI-EXPANDED), (2) Social Sciences Citation Index (SSCI), (3) Arts & Humanities Citation Index (A&HCI), and the (4) Emerging Sources Citation Index (ESCI) were searched. In addition, only articles published in English were considered for analysis. Table 1 lists out the resulting number of papers from the searches after excluding the common papers identified in step 1. A total number of 106 additional papers were identified using the keyword searches, making the total number of papers that had to be reviewed to 198. However, as the major objective of this review is to conduct a meta-analysis, testing the effect of different variables on accuracy levels, studies which did not report any indicator measuring accuracy were not considered for the review. Different studies reported different indicators that showed the accuracy of the methods. Many studies directly mentioned the accuracy percentages (Gu et al 2019a, b), meanwhile, others reported the error percentage in the form of mean

Table 1 Search strategy for literature review

Description of the search	Numbers
Papers from previous review studies ($N=3$) (Do et al. 2019; Nguyen et al. 2018; Wang et al. 2019)	72
Relevant keyword searches in WoS database and number of papers (after accounting for common papers)	
(TS = (deep learning AND traffic flow prediction)) AND LANGUAGE: (English) Indexes = SCI-EXPANDED, SSCI, A&HCI, ESCI Timespan = All years	33
(TS = (deep learning AND traffic speed prediction)) AND LANGUAGE: (English) Indexes = SCI-EXPANDED, SSCI, A&HCI, ESCI Timespan = All years	14
(TS = (deep learning AND traffic accidents)) AND LANGUAGE: (English) Indexes = SCI-EXPANDED, SSCI, A&HCI, ESCI Timespan = All years	9
(TS = (deep learning AND travel demand)) AND LANGUAGE: (English) Indexes = SCI-EXPANDED, SSCI, A&HCI, ESCI Timespan = All years	16
(TS = (deep learning AND transport mode)) AND LANGUAGE: (English) Indexes = SCI-EXPANDED, SSCI, A&HCI, ESCI Timespan = All years	3
(TS = (deep learning AND travel behavio$r)) AND LANGUAGE: (English) Indexes = SCI-EXPANDED, SSCI, A&HCI, ESCI Timespan = All years	1
(TS = (deep learning AND driver behavio$r)) AND LANGUAGE: (English) Indexes = SCI-EXPANDED, SSCI, A&HCI, ESCI Timespan = All years	28
(TS = (deep learning AND distracted driving)) AND LANGUAGE: (English) Indexes = SCI-EXPANDED, SSCI, A&HCI, ESCI Timespan = All years	3
(TS = (deep learning AND travel time prediction)) AND LANGUAGE: (English) Indexes = SCI-EXPANDED, SSCI, A&HCI, ESCI Timespan = All years	6
(TS = (deep learning AND congestion)) AND LANGUAGE: (English) Indexes = SCI-EXPANDED, SSCI, A&HCI, ESCI Timespan = All years	12
(TS = (deep learning AND traffic state prediction)) AND LANGUAGE: (English) Indexes = SCI-EXPANDED, SSCI, A&HCI, ESCI Timespan = All years TS = Topic Search	1
Total studies identified	198
Studies not included and the reasons for their omission	
The study did not mention accuracy, MAPE, MRE, RMSE%, recall, R^2, or AUC values; The studies only considered shallow neural network-based models; used simulated data or the study was not empirical	62
Total studies considered for analysis	136

absolute percentage error (MAPE) (Bao et al. 2019a), mean relative error (MRE) (Zhao et al. 2017), or root mean square error % (RMSE %) (Jo et al. 2019). The error values were used to then calculate the accuracy levels (100 − error%). In addition, some studies also mentioned recall rate (Zhu et al. 2019), R^2 (Polson and Sokolov 2017), and area under curve (AUC) values (Singh and Mohan 2018), which indicated the accuracy levels. Studies which did not report any of the above-mentioned indicators (see Table 1) were omitted from the list of studies to be reviewed. In addition, there were studies which utilised simulated data for their analysis (Gang et al. 2015), they were also removed from the final list of studies to be reviewed. The final tally of papers which were then considered and reviewed for the meta-analysis was 136. The next section describes the review process and the methodology adopted for the meta-analysis.

Meta-analysis

To conduct the meta-analysis, information under nine separate heads were extracted from the papers short-listed for the review ($N=136$). Table 2 lists out the factors extracted. They included recording the (1) year of publication, (2) country where the study was performed or where the data belonged to. (3) The region, i.e. whether an urban area or rural area or both, from where the data were collected. (4) Source of data, (5) the type and format of data used for the analysis, (6) total number of samples considered for the study, (7) the time horizon of prediction, (8) the method used for analysis, and finally (9) the prediction accuracy of the method. The information was collated for all 136 studies which resulted in the collection of 2314 unique rows of information. Each row denoting a method analysed for a particular area of application.

The meta-analysis on the prediction accuracy was carried out using linear mixed effects models (Laird and Ware 1982), which accounted for the random effects representing unobserved heterogeneity in prediction accuracy across studies. Introducing the random effects is essential for the meta-analysis since the accuracy level would depend not only on variables introduced, but also on other unobserved variables such as quality of sensors and data cleaning process.

Table 2 Factors extracted from the literature review and the variables used for meta-analysis

Variable	Description
Factors extracted from the literature	
Year	Year of publication
Coverage	The coverage country of the study
Region	The area or region from where the data was collected, i.e. urban or rural or both
Survey	The source of the data
Data type	The type and the format of the data
Sample size	Total number of samples used in the study generally for training, testing, and validation of the models
Time of prediction	The time of prediction in minutes
Method	The type of method used
Accuracy	The prediction accuracy of the method; calculated through one of accuracy or error indicator
Variables used in the meta-analysis models	
Dependent variable	
Accuracy	The percentage accuracy of a method
Model A	All observations (cases = 2314, studies = 136); all accuracy indicators
Model B	Cases with MAPE; MRE; average accuracy indicators (cases = 1878, studies = 86)
Model C	Cases with accuracy indicator (cases = 220 studies = 29)
Model D	Traffic forecasting studies with sample size and time horizon of prediction information; cases with MAPE; MRE; average accuracy indicators (cases = 991, studies = 36)
Explanatory variables (fixed effects)	
TM	Dummy variable denoting the use of traditional methods; 1 = if yes; else 0
DNN	Dummy variable denoting the use of deep feedforward or backpropagation based neural network; 1 = if yes; else 0
SAE	Dummy variable denoting the use of stacked auto-encoders; 1 = if yes; else 0
DBN	Dummy variable denoting the use of deep belief network; 1 = if yes; else 0
CNN	Dummy variable denoting the use of deep convolutional neural network; 1 = if yes; else 0
LSTM	Dummy variable denoting the use of long short-term memory models; 1 = if yes; else 0
CNNLSTM	Dummy variable denoting the use of combined CNN-LSTM models; 1 = if yes; else 0
Speed	Dummy variable denoting the area of application; 1 = if speed prediction; else 0
Flow	Dummy variable denoting the area of application; 1 = if flow prediction; else 0
Travel demand	Dummy variable denoting the area of application; 1 = if travel demand prediction; else 0
Driver behaviour	Dummy variable denoting the area of application; 1 = if driver behaviour prediction; else 0
TTP	Dummy variable denoting the area of application; 1 = if travel time prediction; else 0
Accident	Dummy variable denoting the area of application; 1 = if accident prediction; else 0
GPS	Dummy variable denoting the type of data used; 1 = if GPS; else 0
Loop detector	Dummy variable denoting the type of data used; 1 = if loop detectors; else 0
CCTV	Dummy variable denoting the type of data used; 1 = if CCTV or camera; else 0
Secondary source	Dummy variable denoting the type of data used; 1 = if secondary sources such as weather information; else 0
Urban	Dummy variable denoting the region of study; 1 = if urban area; else 0
Sample size	The sample size of the study
Time of prediction	The time horizon of prediction in minutes
Explanatory variables (random effects)	
Study	The random effects due to the study (N = 136, 86, 29, and 36 for models A, B, C, and D, respectively)

Percentage accuracy was considered to be the dependent variable. Meanwhile, fixed effects were estimated for different variables denoting the type of method used, area of application, source of data, region of data collection, sample size, and time horizon of prediction.

Four separate models were developed as the information on sample size and time of prediction were not available

for all studies and because different studies used different indicators to measure the prediction accuracy. Model A to C, which did not consider sample size and time of prediction as explanatory variables considered 136, 86, and 29 studies, respectively ($N = 2314$, 1878, and 220, respectively) and tested the effect of other variables listed in Table 2. Model D was developed exclusively for studies which analysed traffic-related applications such as flow and speed and which used the MAPE, MRE, and average accuracy indicators to estimate prediction accuracy ($N = 991$, studies = 36). This model also tested the impact of sample size and time horizon of prediction on the prediction accuracy. The linear mixed effects models were developed in R using the lme4 package (Bates et al. 2014). The results of the meta-analysis are discussed in Sect. 4.

Findings from the Literature Review

As mentioned in the previous section, the information from the papers were extracted and then collated under different heads and this section is dedicated towards describing those factors. It is aimed at providing a detailed overview of the studies, the type of methods, their coverage, and the areas of application. In addition, this section would also discuss the average accuracy observed across different methods and areas of application.

Areas of Application

The review of the literature showed that deep learning methods were used in ten distinct transportation related fields. The most common areas of application belonged in the field of traffic forecasting. Traffic flow forecasting was analysed in 42 studies. Meanwhile, 26 studies used deep learning methods to predict traffic speed (see Table 3). In addition, one study predicted the road occupancy levels (Zang et al.

2017a, b). Travel demand prediction was also observed to be popular with 19 studies estimating the travel demand in different travel modes such as bus (Baek and Sohn 2016), bus rapid transit systems (BRT) (Liu and Chen 2017a), car sharing (Zhu et al. 2017), mass rapid transit systems (MRT) (Liu and Chen 2017b), taxis (Xu et al. 2018a, b, c, d), trains (Tang et al. 2019a, b), for parking (Yang et al. 2019), and bike sharing (Xu et al. 2018a, b, c, d). Moreover, deep learning methods were also used to estimate the travel demand between origins and destinations (Cheng et al. 2017). Prediction of congestion was analysed in 11 different studies. Meanwhile, traffic accidents were predicted in 12 studies, out of which only one study focused on railway accidents (Feng et al. 2018). Meanwhile, others analysed road traffic accidents. Driver behaviour, which included the prediction of distracted driving (Eraqi et al. 2019), subjective risk perception while driving (Ping et al. 2018), lane changing behaviour (Dou et al. 2018), and braking behaviour (Christopoulos et al. 2018) was predicted in 17 studies. Meanwhile, travel behaviour such as mode choice and activity classification was predicted using deep learning methods in five (5) studies. Mode choice was predicted in three studies. Whereas, activity state classification was analysed in one study (Cui et al. 2018a, b, c). One study estimated both mode and activity classification together (Zhao et al. 2019a, b, c, d, e, f). Travel time was predicted in seven (7) studies, which also included one study which predicted both travel time and travel distance together (Jindal et al. 2017) (see Table 3).

Year-Wise Distribution of Studies

The review of the literature clearly showed a steady increase in the use of deep learning methods in the field of transportation. Table 3 shows how the use has increased from merely two (2) studies in 2014 to 59 studies in 2019. Comparison of publications based on the area of application shows a rising

Table 3 Area of application and year of publication

	2014	2015	2016	2017	2018	2019	Total
Accidents	0	0	1	2	4	5	12
Congestion	0	1	1	3	4	2	11
Driver behaviour (DB)	0	0	1	0	4	12	17
Mode and activity choice (M–AC)	0	0	0	1	3	1	5
Travel demand (TD)	0	0	1	6	3	9	19
Travel time prediction (TTP)	0	0	2	0	1	4	7
TTP and distance (Dist.)	0	0	0	1	0	0	1
Traffic flow	2	2	5	6	7	17	39
Traffic speed	0	1	2	7	4	9	23
Speed; flow and occupancy (Occ)	0	0	0	1	0	0	1
Speed and flow	0	0	0	1	0	1	2
Total	2	4	13	28	30	59	136

trend in the fields of accident analysis, driver behaviour prediction, travel time prediction, and traffic state prediction.

Types of Data Source

Loop detectors were observed to be the most commonly used data source, with 55 studies utilising the data obtained from them. Loop detectors are traffic sensors which are usually installed on roads or at toll-stations to detect vehicles. Using the data, one can estimate the flow, speed, and occupancy of the road segments. The frequency at which a typical detector records and relays information varies. Some studies use information recorded every 30 s, while others use 1-min or 5-min interval records. Researchers often aggregate the data to predict the traffic state in the short term. 15-min aggregation was observed to be common among different studies. Data extracted using GPS was observed to be second most common source, with 38 studies using it. Data from mobile phones and vehicle based intelligent transport systems (ITS) both provided GPS information. The information was often in terms of trips made and their trajectories (Bao et al. 2019a; Ma et al. 2015a, b). The same data source was also used to extract information related to speed (Zhao et al. 2019a, b, c, d, e, f), travel time (Petersen et al. 2019), activity state information (Cui et al. 2018a, b, c), and congestion (Chen et al. 2016a, b). GPS is technically an external source of information, whereas mobile phone sensors such as accelerometer, magnetic, gyroscope, barometer (AMGB) can be classified as internal sources. Five studies could be identified which used data from AMGB. They have been utilised in the areas of mode choice prediction (Qin et al. 2018), braking behaviour (Christopoulos et al. 2018), and congestion analysis (Tu et al. 2017). Image-based data collected from CCTV footage or other cameras (including LIDAR sensors) were also commonly used in 22 different studies. Meanwhile, data from mobile phone-based applications

and platforms were used in 10 studies. The data from this source were mostly used to predict travel demand (Ke et al. 2017a, b) and mode and activity states (Zhao et al. 2019a, b, c, d, e, f). Nine studies also used external accident data. Meanwhile, eight (8) studies used the information collected by automated fare collection (AFC) devices fitted in public transport systems (Liu et al. 2019) or parking stations (Yang et al. 2019) (see Table 4). The data collected from AFC devices was used for predicting travel demand in bus and MRT systems (Baek and Sohn 2016; Liu and Chen 2017b). In addition, it was observed that nine studies used certain other sources of data which involved household surveys (Cui et al. 2018a, b, c), toll-based tag data (He et al. 2019), and car based ITS (Jo et al. 2019). Finally, it was observed that in addition to a primary source of data, 40 studies also utilised secondary information such as road network attributes (Zhu et al. 2019), weather information (Xu et al. 2018a, b, c, d), population data (Bao et al. 2019a), and land use data (Baek and Sohn 2016) for the training of their models. It should be noted that many studies utilised more than one data source

Table 4 Type of data sources used in the studies

Data source	Number of studies
Global positioning system (GPS)	38
Loop detectors	55
Accelerometer, magnetic, gyroscope, barometer sensors (AMGB)	5
Closed circuit television (CCTV) cameras	22
Applications	10
Accident data	9
Automated fare collection (AFC)	8
Intelligent transport systems (ITS)	5
Others	4
Secondary sources	40

Table 5 Country and region coverage of the studies

Country	Number of studies
No mention	10
Australia	1
Brazil	1
China	53
Canada	1
Denmark	1
Egypt	1
Germany	1
Greece	1
Hong Kong	1
India	2
Japan	2
Malaysia	1
Morocco	1
Netherlands	2
Norway	1
Palestine	1
Poland	1
South Korea	4
Taiwan	1
Uganda	1
United Kingdom (UK)	11
United States of America (USA)	45
Region	Number of studies
No mention	13
Urban	75
Both	48

for their analysis and therefore, Table 4 has multiple entries for a single study.

Coverage and Region of Studies

The distribution of the studies based on their country of coverage showed that a very high number of studies came from primarily two countries (see Table 5), China (54) and United States of America (USA) (45). A possible reason behind this might be the readily available data in these countries. In addition, it was observed that there were 11 studies which were based in United Kingdom. Meanwhile, four (4) studies utilised data from South Korea for the analysis. Netherlands, Japan, and India showed two studies each, whereas Australia, Canada, Denmark, Egypt, Germany, Greece, Hong Kong, Malaysia, Morocco, Norway, Palestine, Poland, Taiwan, and Uganda each showed one study based on their data. It should be noted that many studies utilised data from more than one country (Eraqi et al. 2019; Qin et al. 2018) and ten studies did not mention any country where the data were based in (Tran et al. 2018).

The information on the region from where the data were collected showed that 75 studies were based on data specifically collected from urban areas. Meanwhile, 48 studies collected data from both urban or rural areas. In cases of loop detector data, unless it was specifically mentioned that the data were only collected from a city or an urban area, it was assumed that the data included information from both urban and rural areas. Moreover, in 13 studies there was no specific mention of the region from where the data were collected.

Accuracy Indicators

The analysis of the studies showed that 13 different type of indicators were used across 136 papers. A majority of them used either the mean absolute error percentage (MAPE) (58 studies) or mean relative error (MRE) (25 studies) indicators. Both MAPE and MRE essentially have the same formulation but the difference is that MAPE is expressed in percentage, whereas MRE is expressed in proportions (see Table. 6). In 2 out of the 58 studies which used MAPE, the studies did not utilise all observations to calculate MAPE. Ke et al. (2017a, b) in their study on estimating travel demand in taxi services, calculated MAPE for values which had a demand intensity of greater than 10, these samples represented the top 4.45% of all samples. In addition, Bao et al. (2019b) utilised the top 5% of the samples with highest values to estimate MAPE. A total of two studies used a symmetric MAPE (sMAPE) indicator to calculate error values. For example, in Xu et al. (2018a, b, c, d), the denominator (see Table 6 for formula) contains additional parameters for predicted values and a constant ($c = 1$ in their study) to avoid a zero denominator. In addition, five studies used an average accuracy indicator

Table 6 Accuracy indicators used in the studies

Measure	Number of studies	Formula (other descriptions)
Accuracy	29	$= \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}$
AUC	3	Area under curve between recall (true positive rate, y-axis) and false positive rate (x-axis)
Precision	2	$= \frac{\text{True positive}}{\text{True positive+false positive}}$
Error rate	1	$= \frac{\text{Number of incorrect predictions}}{\text{Total number of predictions}}$
MAPE	56	$= 100 \times \frac{1}{N} \sum_{i=1}^{N} \frac{|\hat{y}_i - y_i|}{y_i}$
MAPE@10	1	MAPE in the samples with demand intensity greater or equal to 10 (top 4.45% of the largest samples)
MAPE@5%	1	MAPE with top 5% highest value samples
sMAPE	2	$= 100 \times \frac{1}{N} \sum_{i=1}^{N} \frac{|\hat{y}_i - y_i|}{y_i + \hat{y}_i + c}$
MRE	25	$= \frac{1}{N} \sum_{i=1}^{N} \frac{|\hat{y}_i - y_i|}{y_i}$
Average accuracy	5	$= 100 - \text{MAPE or } 1 - \text{MRE}$
R^2	1	Proportion of the variance in the estimated values that is predictable from the observed values
Recall	5	$= \frac{\text{True positive}}{\text{True positive+False negative}}$
RMSE%	3	$= 100 \times \frac{\sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2}}{\bar{y}}$
RMSEP%	1	RMSE% with 70% critical road sections
Average error rate	1	Percentage error (no specific mention of formula or type)

$\hat{y}_i$: predicted values; y_i: observed values; N: number of predictions; c: constant to avoid zero denominator; $\bar{y}$: average of all observed values

obtained from subtracting MAPE or MRE from 100 or 1, respectively.

Other commonly used indicator includes a straightforward accuracy measure (29 studies), which is ratio of number of correct predictions to the total number of predictions (often multiplied by 100 to convert into percentage). This indicator has been mostly used in discrete classification studies (such as driver behaviour) as opposed to predicting a continuous value (like traffic flow). Opposite to the accuracy indicator, one study used the error rate indicator, where they estimated the ratio between number of incorrect predictions to the total number of predictions. Duives et al. (2019) forecasted prediction movements using GPS trajectory data and predicted the movements in the adjoining cells. They estimated the error rates for the 1st, 5th, and 20th prediction of the sequences. Other accuracy indicators in classification studies include recall rate (5 studies), precision (2 studies), and area under the receiver operating characteristics curve (AUC) (3 studies). Meanwhile, a few studies predicting continuous values used root mean square percentage (4 studies) and R^2 (1 study). Since most indicators provide a ratio or proportion roughly denoting the accuracy or error, they were converted to represent an accuracy percentage for the meta-analysis [e.g. accuracy $= 100 \times (1 - \text{MRE})$]. In addition, additional models were created with only (a) MAPE, MRE, and average accuracy indicators and (b) accuracy indicator. The findings of each model were then compared for further discussions.

Deep Learning Methodologies

The review of the literature showed that 136 studies used a total of 2314 methods to predict different transport related variables. The 2314 times these different methods have been tested are also henceforth referred to as cases in this study and they include the use of 11 different groups of deep learning methods along with traditional (TM) and shallow neural network (SNN) methods (see Table 6). Traditional methods in this study are classified as machine learning methods which do not use neural networks. An array of such methods was extensively tested in different studies (primarily to be compared with deep learning models). These methods included autoregressive integrated moving average method (ARIMA), vector autoregression method (VAR), random forest method (RF), support vector machines (SVM), and XGBoost (XGB) among many other methods including variations of the methods mentioned above. As the focus of this study is primarily on understanding the use of deep learning methods, all different traditional methods have been clubbed into one group. Such an aggregation of non-deep learning methods provides a wider scope to evaluate all different deep learning models in the meta-analysis. However, we realise that aggregating all such methods with different characteristics is a major limitation of our study. Similarly, the SNNs also have been grouped together to generate a baseline for the inference. SNNs denote all those neural networks with a shallow architecture. Table 6 shows the different type of methods based on the area of application. A total of 113 cases were estimated using deep neural networks (DNN) based on feed-forward networks. Wang et al. (2019) classified these models as deep multilayer perceptron (MLP) and discussed in detail about their differences with stacked auto-encoders (SAE) and deep belief networks (DBN). SAE and DBN were used to predict 222 and 114 cases, respectively. Recurrent neural networks have been the most popular deep learning method out of all. For the meta-analysis, we divided them into three separate groups: (a) classical recurrent neural networks (RNN), which were used for prediction in 60 cases, (b) long short-term memory (LSTM), which was used the most, in 289 cases, and finally (c) gated recurrent unit (GRU), which was used for prediction in 48 cases throughout the 136 studies. Additionally, in one case, a combination of LSTM- GRU methods was used for the prediction of traffic speed (Gu et al. 2019a, b). Convolutional neural networks (CNN) were the second most popular deep learning method, which was employed to analyse 278 cases. For the ease of analysis, different variants of CNN methods such as Googlenet (Xing et al. 2019), Resnet (Hu et al. 2019), CNN with attention mechanism (Ran et al. 2019a, b), CNN with generative adversarial networks (GAN) (Lee et al. 2019a, b), graph-based convolution (Zhang et al. 2019a, b, c, d) among other methods have been clubbed together. Combined deep learning models also showed prominence across the studies where mostly CNN model was coupled with a type of RNN model. Combination of CNN and LSTM was most common and was used for predicting 48 cases. Meanwhile, combination of CNN and GRU models were used for prediction in 29 cases. Moreover, CNN was also coupled with other classical RNN structures and was used for prediction 10 cases (see Table 7).

Similar trend was also visible in the distribution of the methods based on the area of application. For traffic flow forecasting, DBN (50) and LSTM (77) models were observed to be the most common. However, in case of traffic speed prediction, the use of CNN (101 out of 278) and GRU models were observed to be high (33 out 48 applied in speed prediction). Use of LSTM models were also observed to be common for traffic speed prediction (90 cases). In addition, for travel time prediction, it was observed that DBN models were the mostly used (44 cases). As the prediction of driver behaviour and congestion often used image-based data, the use of CNN was found to be common in these areas of application (38 and 23 respectively). The use of combined deep learning models was most commonly used to predict speed (46), followed by traffic flow (20), travel demand (9), driver behaviour (4), travel time (3), accidents (3), and congestion prediction (2). Finally, Table 6 also lists out the cases estimated using traditional and shallow neural network methods.

Table 7 Deep learning methods and the area of application

	Accidents	Congestion	DB	M–AC	TD	TTP	Dist	Flow	Speed	Occ	Total
CNN	36	23	38	4	20	10	0	43	101	3	278
CNN-GRU	0	0	2	0	0	0	0	9	18	0	29
CNN-LSTM	3	2	2	0	9	3	0	11	18	0	48
CNN-RNN	0	0	0	0	0	0	0	0	10	0	10
LSTM- GRU	0	0	0	0	0	0	0	0	1	0	1
DBN	2	2	5	0	0	44	0	50	10	1	114
DNN	16	1	5	2	10	2	2	38	37	0	113
GRU	0	8	0	0	3	2	0	2	33	0	48
LSTM	10	16	12	4	69	11	0	77	90	0	289
RNN	1	9	12	1	1	4	0	12	20	0	60
SAE	4	1	0	0	14	20	0	151	32	0	222
TM	59	27	46	19	90	35	1	324	218	1	820
SNN	7	14	7	1	42	6	0	142	62	1	282
Total	138	103	129	31	258	137	3	859	650	6	2314

Abbreviations for area of application: DB, driver behaviour; M–AC, mode and activity choice; TD, travel demand; TTP, travel time prediction; Dist., travel distance; Occ, occupancy

Out of 2314, in 820 cases, the variables were predicted using traditional methods. Meanwhile, in 282 cases, they were predicted using SNNs. The next sub-section describes the measures of central tendency in accuracy based on the type of method and the area of application.

Accuracy Distribution Across Different Areas of Application

Figure 1 shows the trends for distribution of accuracy levels for each method and area of application. The distribution is represented with the help of box and whisker diagrams. The boxes represent the mid quartiles, separated by the median value. Meanwhile, the whiskers represent the upper and lower quartiles, whereas the dots in diagrams represent the outliers. Two images are created using different accuracy indicators: the first image (top-image in Fig. 1) was created using MAPE, MRE, and average accuracy indicators and represent the prediction of continuous values. A total of 1878 of out 2314 cases are represented through this image. Meanwhile, the second image (bottom-image in Fig. 1) was created using accuracy indicator and it represents the prediction of discrete states (i.e. classification-based studies). A total of 220 out 2314 cases are represented in that image.

From the plots, it seems that the distribution of the accuracy levels was dependent on the area of application. Accident prediction featured in both images, i.e. they were predicted both as continuous and discrete factors. When predicted as a continuous value, in general, most methods showed a high range of distribution (highest variation in accuracy levels among all areas of application), whereas the median values for accident prediction seemed to be lower than other areas of application. However, when accident was predicted as a discrete factor, then the median value of accuracy seemed to be relatively higher. Meanwhile, accuracy levels for congestion, driver behaviour, and travel demand prediction also showed a relatively higher range of distribution as compared to traffic state prediction variables (see Fig. 1). Traffic state prediction variables such as flow and speed showed a lower range of distribution and higher median values, with usually all methods mostly ranging above 75% accuracy level. In addition, some amount of differences in the distribution with respect to the method applied for prediction were also observed. In the next section, the results of the meta-analysis, discussing the effect of both areas of application and type of method, along with other variables on prediction accuracy are illustrated.

Results

For the meta-analysis, four separate models were developed. The first model (model A) analysed all 2314 samples, testing the effects of deep learning methods, traditional methods, areas of application, type of data source, and the region of study. In addition, the model tested the random effects due to the study ($N = 136$). For this model, all accuracy indicators were converted into a variable with a value out of 100, denoting the level of accuracy. However, as explained earlier, these indicators have different formulations, so two additional models were developed; model B, with only MAPE, MRE, and average accuracy indicators (as they have the same formulation) and model C, with only accuracy indicator. In addition, as the
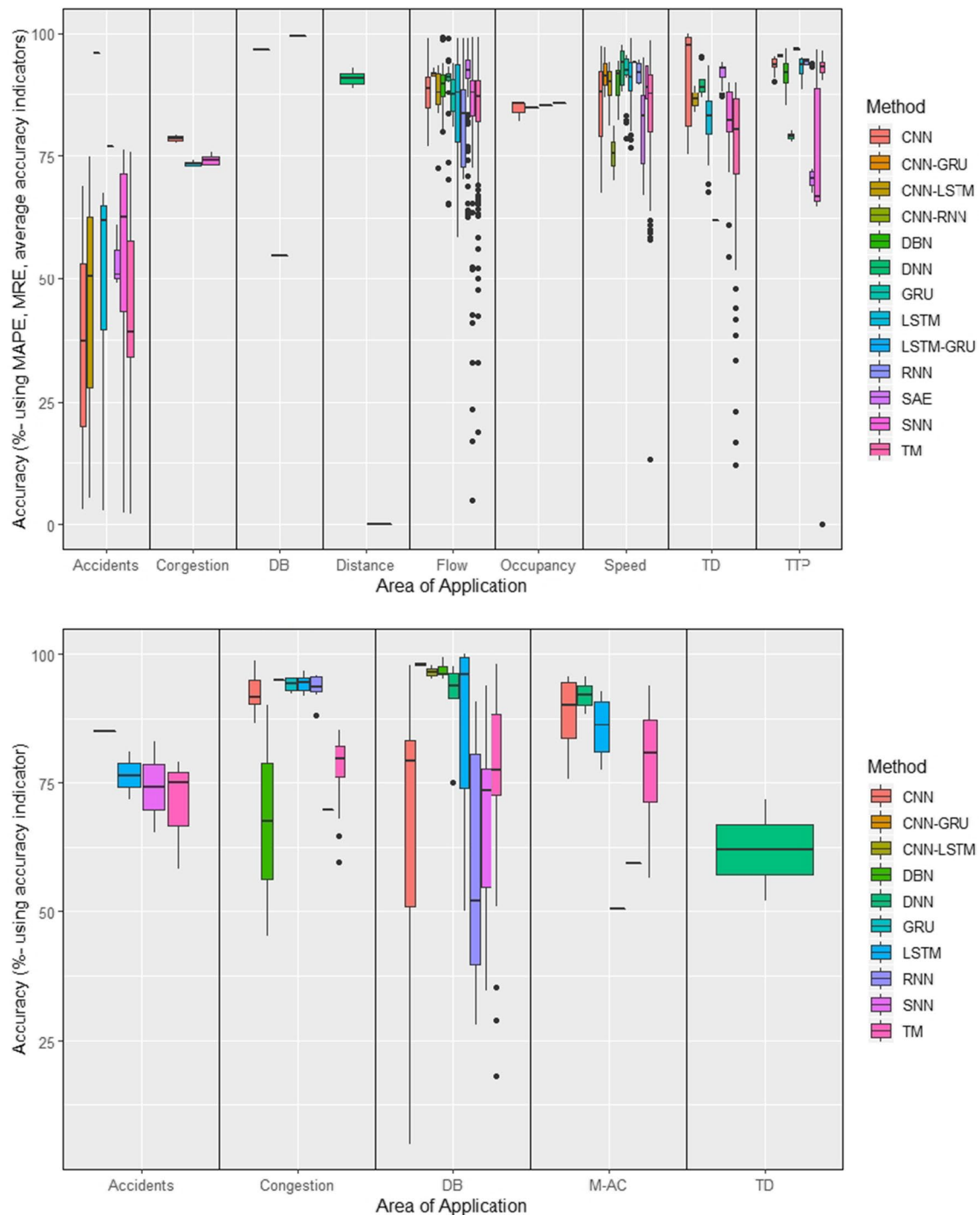
Fig. 1 Accuracy box plot based on area of application and type of method

information on sample size and time horizon of prediction was not available for all the studies, another model (model D) was developed to test the impact of those variables. Model D only considered studies with MAPE, MRE, and average accuracy indicators, which forecasted speed and flow, and contained the information on the sample size and the time horizon of prediction variables. Time horizon of

prediction variable indicates the time in future (generally in minutes) for which a variable is predicted, e.g. predicting the traffic flow in the next 15 min or 30 min or 45 min. This section discusses the results of the models.

Meta-analysis with All Observations and Accuracy Indicators

Random Effects

In this model, one variable representing the random effect due to the study was introduced and study specific intercepts were estimated. The random effect allowed us to account for the unobserved intrinsic heterogeneities among the studies. Table 8 reports the variance of the intercept estimates. It was observed that the random effect of the studies had considerable variance and the heterogeneities among them have a major contribution towards influencing prediction accuracy. Moreover, the variance in the random effect parameter was observed to be higher than the residual variance. In addition, the comparison of marginal R^2 value, which are associated with the fixed effects and the conditional R^2 value, which are associated with both fixed and random effects showed that the contribution of the random effects was much higher (0.175 vs. 0.780). The result of the likelihood ratio test for the study-level random component also show its significant contribution ($\chi^2 = 1847.60^{***}$).

Table 8 Results of the meta-analysis model for all samples

Models	A		B		C	
	Estimate	t value	Estimate	t value	Estimate	t value
Fixed effects						
(Intercept)	83.95	21.88	79.8	17.59	75.83	8.82
Methodologies						
TM	− 1.72	− 3.11	− 1.81	− 3.64	− 7.99	− 3.73
DNN	3.95	3.5	4.41	4.1	0.08	0.02
SAE	3.09	3.93	3.04	4.56	−	−
DBN	4.06	3.18	3.42	3.06	6.97	1.38
CNN	3.8	4.84	3.84	5.13	1.8	0.75
LSTM	5.07	6.7	4.8	7.04	5.6	1.79
CNN-LSTM	5.52	3.86	4.97	3.78	1.58	0.21
Area of application						
Speed	8.47	2.87	9.45	3.1	−	−
Flow	2.83	0.94	3.63	1.19	−	−
Travel demand	− 8.46	− 1.84	1.72	0.34	−	−
Driver behaviour	− 13.88	− 2.68	−	−	− 9.94	− 1.13
TTP	2.1	0.45	1.05	0.26	−	−
Accident	− 15.51	− 2.94	− 24.33	− 3.72	−	−
Source of data						
GPS	− 8.28	− 2.55	− 4.17	− 1.13	− 1.95	− 0.19
Loop detector	− 2.47	− 0.73	1.47	0.39	−	−
CCTV	3.34	0.82	5.73	1.2	5.12	0.47
Secondary source	4.74	2.91	4.67	2.99	5.63	0.62
Target area						
Urban	− 3.23	− 3.66	− 3.6	− 4.86	8.03	0.92
Random effects (variance)						
Study	189.53		115.83		366.05	
Residual	68.77		46.82		86.02	
Model performance parameters						
Sample size (cases)	2314		1878		220	
Sample size (study)	136		86		29	
AIC	16,781.89		12,828.8		1649.86	
Log likelihood	− 8369.94		− 6394.401		− 810.9298	
R^2 (only with fixed effects)	0.175		0.212		0.123	
R^2 (with fixed and random effects)	0.78		0.773		0.833	

Model A: with all observations and indicators; Model B: with only MRE, MAPE, and average accuracy indicators; Model C: with only accuracy indicator

Fixed Effects: Methodologies

For methodologies, seven dummy variables (6 for deep learning methods and 1 for traditional methods) were tested as predictor variables in the linear mixed effects model. It was observed that all deep learning methods showed a significant positive relationship with the prediction accuracy. A comparison of parameter estimates across different methods showed that the combined CNN-LSTM had the largest significant positive effect on prediction accuracy, followed by LSTM and DBN models. In addition, DNN, CNN, and SAE models also showed significant positive effects on the accuracy levels (see Table 8). Meanwhile, the effect of traditional methods was observed to be significantly negative. The findings clearly established that deep learning methods produce estimates with better prediction accuracies when compared to traditional methods. In addition, the findings indicate that combined CNN-LSTM models might be better as compared to other models when it comes to prediction accuracies. However, it should be noted that not all models can be applied to predict all variables. The combined CNN-LSTM models have been most commonly applied in the prediction of flow and speed. However, the accuracy levels produced from their application in the field of accident and travel demand prediction showed a high range of distribution (see Fig. 1). The intrinsic properties of a particular model make them suitable for certain applications. For example, it is known that traffic flow and speed have strong spatial and temporal dependencies, i.e. the current traffic flow or speed depends largely on the previous and surrounding flow or speed conditions. In such a case, the combined CNN and RNN-type models would perform well due to its ability to handle both spatial and temporal dependencies. Another possible reason behind the high positive significance of the deep learning models might be linked with researchers using them for the right purpose and area of application. Having a sound knowledge about the nature of the data and applying the most appropriate model for prediction could be possible reasons for the positive relationships. In addition, it should be noted that the papers reviewed in this study were all deep learning-based papers, often proposing the application of new deep learning models. It is possible that when compared with other papers which incorporated and tested only traditional machine learning, the results of the meta-analysis might change. Nevertheless, this study provided an empirical basis to evaluate the significance and contribution of deep learning methods towards improving the prediction accuracy.

Fixed Effects: Area of Application

The effect of six dummy variables representing the areas of application was tested. It was observed that analysing and predicting traffic speed positively affected the prediction accuracy. Meanwhile, predicting travel demand, driver behaviour, and accidents had a significant negative impact on accuracy levels. A comparison of the parameter estimates showed that the dummy variable for accident prediction had the strongest impact, followed by the variable for driver behaviour (see Table 8). Other two variables, traffic flow and travel time prediction, showed positive relationships, but the estimates were not statistically significant.

Fixed Effects: Other Variables

Apart from the type of method and the area of application, dummy variables denoting the source of data and the region of study were also used as predictor variables in the model. The relationship of the dummy variable signifying the use of GPS data was observed to be negative. It should be noted that these results are for all methods and cases. A possible reason behind this might be the nature of errors in the GPS data. Unlike other data sources such as the loop-detectors, it is possible that the errors in GPS data are more random and difficult to model. Using a secondary source to train parameters such as weather data and road network attributes. showed a significant positive relationship with accuracy levels. Meanwhile, other variables denoting the use of loop detector data and image-based data from cameras were not found to be statistically significant (see Table 8). In addition, it was observed that studies which were conducted in an urban area had a significant negative relationship with prediction accuracy. This finding could be intuitively understood as data from urban areas often have a lot of complexities and therefore, the entanglement of these complex factors of variation makes it more challenging to produce a higher prediction accuracy.

Meta-analysis with Selected Indicators

Random effects

Similar to the model with all observations, for both model B and C, i.e models with observations using MAPE, MRE, or average accuracy indicators (all converted to denote accuracy out of 100) and accuracy indicator, respectively showed that random effect of the studies had considerable variance and the heterogeneities among them have a major contribution towards influencing prediction accuracy. In addition, the comparison of marginal R^2 value, which are associated with the fixed effects and the conditional R^2 value, which are associated with both fixed and random effects showed

that the contribution of the random effects was much higher for both the models (0.212 vs. 0.773 in model B and 0.123 vs. 0.833 in model C). The result of the likelihood ratio test for the study-level random component also show its significant contribution ($\chi^2 = 1204.00^{***}$ for model B and $\chi^2 = 166.57^{***}$ for model C).

Fixed Effects: Model B

In model B, the findings of the effect of deep learning methodologies were observed to be exactly similar to the findings of model A. It was observed that deep learning methods had a significant positive effect on prediction accuracy, whereas traditional machine learning methods had a negative effect. In addition, similar to model A, the dummy variable for CNN-LSTM method showed the strongest positive effect among all variables for deep learning methods. In the case variables denoting areas of application, 'speed' showed a positive relationship with accuracy. Meanwhile, 'accident' showed a negative relationship (see Table 8). The effect of other variables was not observed to statistically significant. In addition, in the case of variables denoting data source, only the dummy variable indicating the use of secondary data source showed a statistically significant, positive relationship with prediction accuracy. The effect of other dummy variables denoting data source was not observed to be statistically significant. Finally, the effect of region of study, i.e. whether the study was conducted in an urban area or not, showed a significant negative relationship with prediction accuracy, this finding too was similar to model A.

Fixed Effects: Model C

Many variables tested in models A and B could not be tested in model C because of the low sample size, as model C only considered classification-based studies which used the accuracy indicator (29 studies). For the effect of methodologies, only the variables for LSTM and traditional methods showed statistically significant relationships (see Table 8). The findings were on expected lines, LSTM showed a positive relationship, whereas traditional methods showed a negative relationship. Among the variables denoting the areas of application, only the effect of driver behaviour was tested, and the result was not statistically significant. In addition, the effect of variables denoting data source and target area was also not statistically significant.

Table 9 Results of the meta-analysis model for traffic flow and speed studies

	Estimate	t value
Intercept	85.06	29.17
Methodologies		
TM	− 1.58	− 3.54
DNN	1.99	2.45
SAE	1.72	2.49
DBN	2.97	2.98
CNN	3.38	5.21
LSTM	2.98	4.89
CNN-LSTM	4.5	3.59
Loop detector	3.66	3.76
Sample size (1E−8)	5.9	2.46
Time horizon of prediction (min) (1E−2)	− 5.28	− 8.03
Random effects (variance)		
Study	244.25	
Residual	22.76	
Model performance parameters		
Sample size (cases)	991	
Sample size (study)	36	
AIC	6105.631	
Log-likelihood	− 3039.816	
R^2 (only with fixed effects)	0.485	
R^2 (with fixed and random effects)	0.956	

Meta-analysis for Traffic Flow and Speed Studies

Random Effects

Similar to the other linear mixed effects models (A–C), the model with a sub-sample including studies which predicted traffic speed and flow and studies which used MAPE, MRE, and average accuracy indicators (model D) also showed high variance. The sub-sample, which contained the information on both the sample size and the time horizon of prediction included a total of 991 observation across 36 studies. In addition, it was also observed that the variance in the random effect parameter was higher than the residual variance (see Table 9). In this model also, the R^2 is drastically improved after adding random effects (improved from 0.485 to 0.956), which is indicative towards the high contribution of the random effects in explaining the total variance. The result of the likelihood ratio test also supported the same ($\chi^2 = 1098.80^{***}$).

Fixed Effects: Sample Size and Time of Prediction

Total sample size, i.e. the number of observations used for training, testing, and validation was used as an explanatory variable and its impact on the prediction accuracy was

estimated. It was observed that sample size had a significant positive relationship with the prediction accuracy (see Table 9). Although expected, but it is an important finding with respect to the future of deep learning and its application in transport studies. This finding can be interpreted to mean that to better predict traffic speed or flow, the models would require a large sample size. This would mean the requirement of more resources in (1) financial aspects and (2) computational aspects, to acquire and handle the big data available from various sources.

Time horizon of prediction variable, as explained earlier is the time in future for which the variable is estimated and it was observed that it had a significant negative relationship with the prediction accuracy. This means that as the time horizon of prediction increases, the prediction accuracy of traffic forecasting decreases. Predicting long-term (few hours to weeks) traffic estimates have always been identified as challenging and are prone to errors (Jiang and Adeli 2005). However, prediction accuracies in the period of 60–120 min also need to be improved. Accurate traffic prediction for longer terms can prove to be an important aspect of transportation planning and management. Better prediction accuracies across different time horizons would increase the possibility of introducing challenging but efficient policies such as dynamic road pricing by providing users with the authority to make more flexible and sustainable transport choices.

Fixed Effects: Other Variables

The meta-analysis model using the sub-sample of traffic forecasting studies also tested the effects of methodologies and data source types. The findings were observed to be similar to that of the earlier models (A and B). It was observed that deep learning methods showed significant positive effects on prediction accuracy, with the combined CNN-LSTM model having the strongest relationship with traffic forecasting accuracy (see Table 9). Meanwhile, the relationship of traditional methods was observed to be significantly negative. The effect of data source type by creating a dummy variable for loop detectors was also tested. It was observed that it had a significant positive relationship with accuracy.

Conclusions

This review paper conducted a comprehensive survey of the application of deep learning methods in transport studies. Following a detailed search strategy, a total 136 studies were selected. These studies were reviewed, and information was extracted for several important variables to test their relationship with prediction accuracy.

Before this study, three more review studies had concisely summarised the type of methodologies and the areas of application. This study extends these previous works and adds new knowledge in the following ways: First, as the application of deep learning in transportation is fast growing, many new studies were not reviewed in the previous review papers were added and reviewed. Second, we analysed the papers dealing with the application of deep learning with respect to various other variables which were not looked into earlier, including the coverage and region of the study, the type of data source, sample size, time horizon of prediction, and the prediction accuracies for different methods and areas of application. Finally, by conducting a meta-analysis we could empirically establish the relationships between influential factors and the prediction accuracy. In this section, we would like to discuss and summarise our key findings with respect to the future of deep learning and AI in transportation. The summary is presented in relation to the most relevant factors analysed in the study. Figure 2 presents a graphical summary of the survey, representing the type of deep learning methods, data sources, and areas of applications used across the 136 studies. In addition, it shows the connections between these variables vis-à-vis prediction accuracy.

Prediction accuracy The meta-analysis considered accuracy levels as the dependent variable and tested its relationship with different variables. High prediction accuracy might be one of the most important reasons behind the growing popularity of the deep learning methods. Moreover, obtaining a high rate of prediction accuracy would also ensure the formulation and implementation of successful transportation policies based on future forecasting. People's use of and dependence on information and communication technologies (ICT) have been continuously increasing. The predictive powers of our algorithms should be high enough to correctly estimate the interactions between ICT systems, transportation market, and the higher order impacts on land use, congestion levels, changes in emission levels, etc. AI would somehow be at the centre of all this. However, this analysis clearly established that not all methods are equal, though the required accuracy level for practical use would also depend on areas of application. In addition, it was observed that there existed high variance due to the heterogeneities intrinsic to a study and the accuracy significantly depended on many other predictor variables. The challenge in front of future researchers and policy-makers would be to identify and control these factors.

Methodologies The results of the analysis showed the differences among the methodologies. Combined CNN-LSTM models showed the strongest positive influence. Meanwhile, other deep learning models also showed positive effects. CNN-LSTM models are an improvement over certain deep learning methods as the combination of both methods
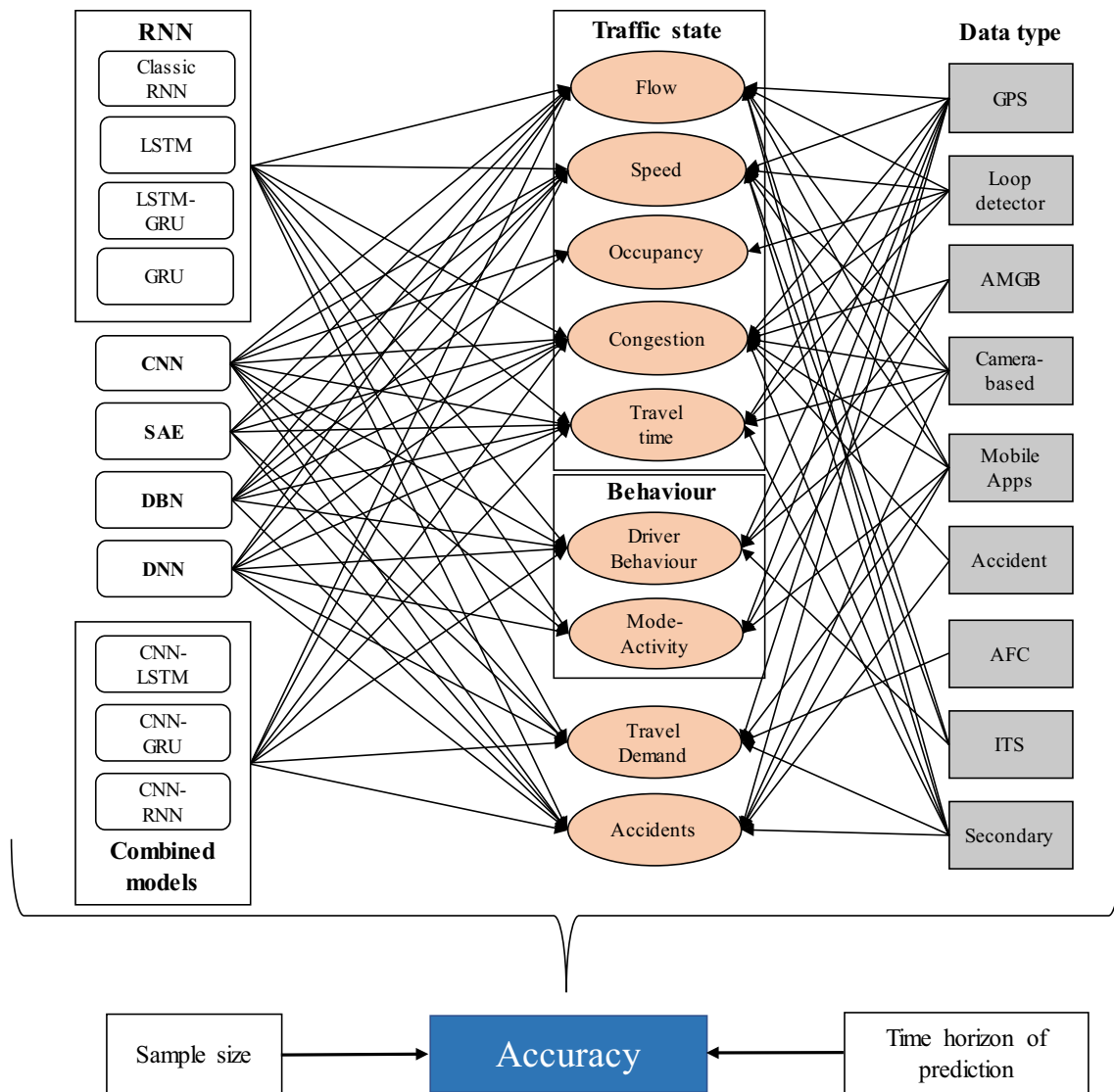
Fig. 2 Deep learning in transportation: a graphical summary

enables to extract spatiotemporal features and correlations (Yu et al. 2017a, b, c). CNN is utilised to understand the spatial dependencies. Meanwhile, LSTM is employed on the time axis to understand temporal dependencies. Currently, a high percentage of this method's application has been witnessed in traffic speed and flow forecasting, but it has also been applied to other areas. Accidents, travel demand, and congestion can also be predicted using the same. However, the variation in the accuracy levels was observed to be high in other areas of application. Another possible drawback to this method can be related tp RNN-based networks being difficult to train and requiring high computational power. Studies such as by Yu et al. (2017a, b, c) have tried to overcome these issues by employing a fully convolutional structure on the time axis and developing spatiotemporal CNNs. For this analysis, spatiotemporal CNNs have been clubbed

with other CNN models and their effects were also observed to be positive. Ultimately, it would be upon policy-makers and researchers to make a trade-off between training time, computational requirement, and prediction accuracy. The decision of which type of model to select would also depend upon its area of application.

Areas of application and data sources Ten distinct areas of application could be identified from the survey. The major focus has been on the prediction of traffic speed and flow, with 1509 out of 2314 cases being from these two areas. A possible reason behind this might be the available loop detector and GPS data, which has been mostly used in traffic state prediction (see Fig. 2). However, certain studies also utilised information obtained from car-ITS (Xiangxue et al. 2019) and camera-based sources (He et al. 2019) to predict traffic parameters. The meta-analysis model showed

the significance of four areas of application variables: traffic speed, travel demand, driver behaviour, and accidents. Given the high amount of data available for traffic forecasting (i.e. speed and flow) and the amount of work conducted in this field (Do et al. 2019; Wang et al. 2019), the positive effect was an expected finding. The challenge currently is to improve the performance in areas where the models have showed a poor performance, e.g. travel demand, driver behaviour, and accident prediction. Travel demand studies utilised data from GPS, mobile applications, and AFC data sources. Meanwhile, driver behaviour studies have utilised four types of data sources, GPS, camera based, AMGB, and ITS based. In addition, it was also observed that accident prediction utilised GPS, camera-based, and secondary accident data to develop deep learning architectures. These areas of application are important aspects of transportation with pertinent policy implications. For example, driver behaviour is an important area of application with respect to the discussions around automated vehicles (AV). AVs can be classified into five levels of automation and accurately predicting driver behaviour would be very important in cases of partial and conditional automation. There has been immense focus on the image-based classification of vehicles, traffic signs, and pedestrians. However, improving the accuracy levels for driver behaviour prediction would be an important task for future.

The other areas of application (driver behaviour and accident prediction) also need further improvement as their relationships with accuracy level were observed to be negative. Data from ITS such as AFC systems (Tang et al. 2019a, b) have provided an opportunity to predict travel demand in public transport modes. Accurate predictions of demand would be beneficial in future policy-making such as information provision on crowding inside public transport. Moreover, it might be useful in the application of mobility as a service (MaaS) schemes. However, a major challenge in the field of transportation that remains is the black-box nature of deep learning methods and the difficulty in their interpretation (Wang et al. 2019; Zhang and Zhu 2018). Travel behaviour models for mode choice and activity participation and traffic flow models are grounded in theory and a major challenge for future would be to improve the interpretability of the deep learning models. In this regard, for example, a discrete choice model with neural network elements can be developed without the compromise of behavioural interpretability by adding a certain constraint on the parameters that need to be behaviourally explained (Sifringer et al. 2018).

Sample size and time of prediction There was no unique way found in the literature to report sample sizes. Data collected from loop detectors often reported the number of days for which and the number of loop detectors from where the data were collected (Tian et al. 2018a, b). In addition, they also reported the frequency of aggregation (e.g.

1 min, 5 min, or 15 min). Meanwhile, data from GPS sources reported trip trajectory information which was then utilised to extract speed, flow, travel demand data, or other information based on the area of application (Bao et al. 2019a; Elhenawy and Rakha 2017; Zhu and Laptev 2017). Contrasting to these sources, image-based sources such as CCTV and other cameras often had a lower sample size but was utilised to extract multiple features in a single image (Eraqi et al. 2019). Given the difference in scales of the data, it might be difficult to compare across studies. In addition, many studies did not report on the sample size. However, the meta-analysis showed a significant positive effect of sample size on prediction accuracy. Higher sample size would also mean the requirements of higher computational abilities and the challenge for future researches would be to find the correct balance between size of data and the required prediction accuracy. Finally, most studies that have been analysed have predicted the short-term traffic forecast but the time period of prediction for short term prediction also varies. The meta-analysis established that as the time of prediction increases, the prediction accuracy tends to decrease. Accurate predictions which are 30 min to 2 h in advance can prove to be beneficial for both policy-makers and individuals for their personal trip planning. Dynamic congestion pricing, MaaS, traffic state prediction during disruptions and big events are possible areas of application which might be benefitted from improving this aspect.

Along with the key findings, contributions, and discussions for future research, it is important to discuss the limitations of this study. This study did not consider the work done in image-based classification of vehicles, traffic signs, pedestrians, and pavement crack detection. These areas of applications have been comprehensively covered in Wang et al. (2019). Rather it focused on travel-related and travel behaviour factors. Moreover, studies which used deep learning in traffic signal control were not included as they did not usually involve the prediction of any factor. These are important areas of application and not considering them remains a major limitation of this work. Moreover, the papers considered for this research are limited by the search strategy and there is a possibility that relevant work conducted in this field were not included. In addition, the meta-analysis combines the work done in both discrete state prediction (driver behaviour, activity state, etc.) and continuous value's prediction (traffic flow and speed), while combining the different accuracy indicators across these studies. The combination across different scales was primarily done for the ease of analysis and comparison. Finally, hyper-parameters related to deep learning models such as the number of hidden layers and epochs were not considered as part of this analysis.

travel demand prediction and comprehensive transport demand management", supported by the Committee on Advanced Road Technology under the authority of the Ministry of Land, Infrastructure, Transport, and Tourism in Japan.

References

Aqib M, Mehmood R, Alzahrani A, Katib I, Albeshri A, Altowaijri SM (2019a) Rapid transit systems: smarter urban planning using big data, in-memory computing, deep learning, and GPUs. Sustainability 11:2736

Aqib M, Mehmood R, Alzahrani A, Katib I, Albeshri A, Altowaijri SM (2019b) Smarter traffic prediction using big data, in-memory computing, deep learning and GPUs. Sensors 19:2206

BaekSohn SMK (2016) Deep-learning architectures to forecast bus ridership at the stop and stop-to-stop levels for dense and crowded bus networks. Appl Artif Intell 30:861–885

Bai J, Chen Y (2019) A deep neural network based on classification of traffic volume for short-term forecasting. Math Probl Eng 2019:1–10

Bao J, Liu P, Ukkusuri SV (2019a) A spatiotemporal deep learning approach for citywide short-term crash risk prediction with multi-source data. Accid Anal Prev 122:239–254. https://doi.org/10.1016/j.aap.2018.10.015

Bao J, Yu H, Wu J (2019b) Short-term FFBS demand prediction with multi-source data in a hybrid deep learning framework. IET Intell Transp Syst 13(9):1340–1347

Bates D, Mächler M, Bolker B, Walker S (2014) Fitting linear mixed-effects models using lme4. arXiv1406.5823.

Bengio Y, Lamblin P, Popovici D, Larochelle H (2007) Greedy layer-wise training of deep networks. In: Advances in neural information processing systems, pp 153–160

Bin Y, Zhongzhen Y, Baozhen Y (2006) Bus arrival time prediction using support vector machines. J Intell Transp Syst 10:151–158

Celaya-Padilla JM, Galván-Tejada CE, Lozano-Aguilar JSA, Zanella-Calzada LA, Luna-García H, Galván-Tejada JI, Gamboa-Rosales NK, Velez Rodriguez A, Gamboa-Rosales H (2019) "Texting & Driving" detection using deep convolutional neural networks. Appl Sci 9:2962

Chakraborty P, Adu-Gyamfi YO, Poddar S, Ahsani V, Sharma A, Sarkar S (2018) Traffic congestion detection from camera images using deep convolution neural networks. Transp Res Rec 2672(45):222–231

Chen X, Xiang S, Liu C-L, Pan C-H (2014) Vehicle detection in satellite images by hybrid deep convolutional neural networks. IEEE Geosci Remote Sens Lett 11:1797–1801

Chen Q, Song X, Yamada H, Shibasaki R (2016a) Learning deep representation from big and heterogeneous data for traffic accident inference. In: AAAI, pp 338–344

Chen Y, Lv Y, Li Z, Wang F-Y (2016b) Long short-term memory model for traffic congestion prediction with online open data. In: 2016 IEEE 19th international conference on intelligent transportation systems (ITSC). IEEE, pp 132–137

Chen W, An J, Li R, Fu L, Xie G, Bhuiyan MZA, Li K (2018a) A novel fuzzy deep-learning approach to traffic flow prediction with uncertain spatial–temporal data features. Future Gener Comput Syst 89:78–88

Chen M, Yu X, Liu Y (2018b) PCNN: deep convolutional networks for short-term traffic congestion prediction. IEEE Trans Intell Transp Syst 19(11):3550–3559

Cheng Q, Liu Y, Wei W, Liu Z (2017) Analysis and forecasting of the day-to-day travel demand variations for large-scale transportation networks: a deep learning approach. In: Transportation research board 96th annual meeting

Chikaraishi M, Garg P, Varghese V, Yoshizoe K, Urata J, Shiomi Y, Watanabe R (2020) On the possibility of short-term traffic prediction during disaster with machine learning approaches: an exploratory analysis. Transp Policy 98:91–104

Christopoulos SRG, Kanarachos S, Chroneos A (2018) Learning driver braking behavior using smartphones, neural networks and the sliding correlation coefficient: road anomaly case study. IEEE Trans Intell Transp Syst 20(1):65–74

Chung J, Gulcehre C, Cho K, Bengio Y (2014) Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv1412.3555

CireşAn D, Meier U, Masci J, Schmidhuber J (2012) Multi-column deep neural network for traffic sign classification. Neural Netw 32:333–338

Cui Y, He Q, Khani A (2018a) Travel behavior classification: an approach with social network and deep learning. Transp Res Rec 0361198118772723

Cui Z, Henrickson K, Ke R, Wang Y (2018b) High-order graph convolutional recurrent neural network: a deep learning framework for network-scale traffic learning and forecasting. arXiv1802.07007

Cui Z, Ke R, Wang Y (2018c) Deep bidirectional and unidirectional LSTM recurrent neural network for network-wide traffic speed prediction. arXiv1801.02143

Dabiri S, Heaslip K (2018) Inferring transportation modes from GPS trajectories using a convolutional neural network. Transp Res Part C Emerg Technol 86:360–371. https://doi.org/10.1016/j.trc.2017.11.021

Deng S, Jia S, Chen J (2019) Exploring spatial–temporal relations via deep convolutional neural networks for traffic flow prediction with incomplete data. Appl Soft Comput 78:712–721

Do LNN, Taherifar N, Vu HL (2019) Survey of neural network-based models for short-term traffic state prediction. Wiley Interdiscip Rev Data Min Knowl Discov 9:e1285

Dong W, Li J, Yao R, Li C, Yuan T, Wang L (2016) Characterizing driving styles with deep learning. arXiv1607.03611

Dong C, Shao C, Clarke DB, Nambisan SS (2018) An innovative approach for traffic crash estimation and prediction on accommodating unobserved heterogeneities. Transp Res Part B Methodol 118:407–428. https://doi.org/10.1016/j.trb.2018.10.020

Dou Y, Fang Y, Hu C, Zheng R, Yan F (2018) Gated branch neural network for mandatory lane changing suggestion at the on-ramps of highway. IET Intell Transp Syst 13:48–54

Duan Y, Lv Y, Wang F-Y (2016a) Travel time prediction with LSTM neural network. In: 2016 IEEE 19th international conference on intelligent transportation systems (ITSC). IEEE, pp 1053–1058

Duan Y, Lv Y, Wang F-Y (2016b) Performance evaluation of the deep learning approach for traffic flow prediction at different times. In: 2016 IEEE international conference on service operations and logistics, and informatics (SOLI). IEEE, pp 223–227

Duives DC, Wang G, Kim J (2019) Forecasting pedestrian movements using recurrent neural networks: an application of crowd monitoring data. Sensors 19:382

Elhenawy M, Rakha H (2017) Spatiotemporal traffic state prediction based on discriminatively pre-trained deep neural networks. Adv Sci Technol Eng Syst J 2:678–686

Eraqi HM, Abouelnaga Y, Saad MH, Moustafa MN (2019) Driver distraction identification with an ensemble of convolutional neural networks. J Adv Transp 2019:4125865. https://doi.org/10.1155/2019/4125865

European Parliamentary Research Service (2019) EU guidelines on ethics in artificial intelligence: context and implementation

Fan Z, Liu C, Cai D, Yue S (2019) Research on black spot identification of safety in urban traffic accidents based on machine learning method. Saf Sci 118:607–616

Fang S-H, Fei Y-X, Xu Z, Tsao Y (2017) Learning transportation modes from smartphone sensors based on deep neural network. IEEE Sens J 17:6111–6118

Feng F, Li W, Jiang Q (2018) Railway traffic accident forecast based on an optimized deep auto-encoder. Promet-Traffic Transp 30:379–394

Gang X, Kang W, Wang F, Zhu F, Lv Y, Dong X, Riekki J, Pirttikangas S (2015) Continuous travel time prediction for transit signal priority based on a deep network. In: 2015 IEEE 18th international conference on intelligent transportation systems. IEEE, pp 523–528

Goodfellow I, Bengio Y, Courville A, Bengio Y (2016) Deep learning. MIT Press, Cambridge

Graves A, Mohamed A, Hinton G (2013) Speech recognition with deep recurrent neural networks. In: 2013 IEEE international conference on acoustics, speech and signal processing. IEEE, pp 6645–6649

Gu Y, Lu W, Qin L, Li M, Shao Z (2019a) Short-term prediction of lane-level traffic speeds: a fusion deep learning model. Transp Res Part C Emerg Technol 106:1–16

Gu Y, Shao Z, Qin L, Lu W, Li M (2019b) A deep learning framework for cycling maneuvers classification. IEEE Access 7:28799–28809. https://doi.org/10.1109/ACCESS.2019.2898852

Han D, Chen J, Sun J (2019a) A parallel spatiotemporal deep learning network for highway traffic flow forecasting. Int J Distrib Sens Netw 15:1550147719832792

Han Y, Wang S, Ren Y, Wang C, Gao P, Chen G (2019b) Predicting station-level short-term passenger flow in a citywide metro network using spatiotemporal graph convolutional neural networks. ISPRS Int J Geo-Inf 8:243

He Z, Chow C, Zhang J (2019) STANN: A Spatio-Temporal Attentive Neural Network for Traffic Prediction. IEEE Access 7:4795–4806. https://doi.org/10.1109/ACCESS.2018.2888561

Hochreiter S, Schmidhuber J (1997) LSTM can solve hard long time lag problems. In: Advances in neural information processing systems, pp 473–479

Hong H, Huang W, Song G, Xie K (2014) Metric-based multi-task grouping neural network for traffic flow forecasting. In: International symposium on neural networks. Springer, pp 499–507

Hu Y, Lu M, Lu X (2019) Driving behaviour recognition from still images by using multi-stream fusion CNN. Mach Vis Appl 30:851–865

Huang W, Song G, Hong H, Xie K (2014) Deep architecture for traffic flow prediction: deep belief networks with multitask learning. IEEE Trans Intell Transp Syst 15:2191–2201

Huang X, Sun J, Sun J (2018) A car-following model considering asymmetric driving behavior based on long short-term memory neural networks. Transp Res Part C Emerg Technol 95:346–362. https://doi.org/10.1016/j.trc.2018.07.022

Huang Z, Li Q, Li F, Xia J (2019) A novel bus-dispatching model based on passenger flow and arrival time prediction. IEEE Access 7:106453–106465

Jia Y, Wu J, Du Y (2016) Traffic speed prediction using deep learning method. In: Intelligent transportation systems (ITSC), 2016 IEEE 19th international conference on. IEEE, pp 1217–1222

Jia Y, Wu J, Ben-Akiva M, Seshadri R, Du Y (2017a) Rainfall-integrated traffic speed prediction using deep learning method. IET Intell Transp Syst 11:531–536

Jia Y, Wu J, Xu M (2017b) Traffic flow prediction with rainfall impact using a deep learning method. J Adv Transp 2017:6575947. https://doi.org/10.1155/2017/6575947

Jiang X, Adeli H (2005) Dynamic wavelet neural network model for traffic flow forecasting. J Transp Eng 131:771–779

Jiang J, Lin F, Fan J, Lv H, Wu J (2019) A destination prediction network based on spatiotemporal data for bike-sharing. Complexity 2019:7643905. https://doi.org/10.1155/2019/7643905

Jindal I, Chen X, Nokleby M, Ye J (2017) A unified neural network approach for estimating travel time and distance for a taxi trip. arXiv1710.04350

Jo D, Yu B, Jeon H, Sohn K (2019) Image-to-image learning to predict traffic speeds by considering area-wide spatio-temporal dependencies. IEEE Trans Veh Technol 68:1188–1197

Kanestrøm PØ (2017) Traffic flow forecasting with deep learning. Master thesis, Norwegian University of Science and Technology

Ke J, Zheng H, Yang H, Chen X (2017a) Short-term forecasting of passenger demand under on-demand ride services: a spatio-temporal deep learning approach. Transp Res Part C Emerg Technol 85:591–608. https://doi.org/10.1016/j.trc.2017.10.016

Ke J, Zheng H, Yang H, Chen X (2017b) Short-term forecasting of passenger demand under on-demand ride services: a spatio-temporal deep learning approach. Transp Res Part C Emerg Technol 85:591–608. https://doi.org/10.1016/j.trc.2017.10.016

Kong F, Li J, Jiang B, Song H (2019a) Short-term traffic flow prediction in smart multimedia system for internet of vehicles based on deep belief network. Future Gener Comput Syst 93:460–472

Kong F, Li J, Jiang B, Zhang T, Song H (2019b) Big data-driven machine learning-enabled traffic flow prediction. Trans Emerg Telecommun Technol 30:e3482

Laird NM, Ware JH (1982) Random-effects models for longitudinal data. Biometrics 38:963–974

LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436

Lee S, Ngoduy D, Keyvan-Ekbatani M (2019a) Integrated deep learning and stochastic car-following model for traffic dynamics on multi-lane freeways. Transp Res Part C Emerg Technol 106:360–377

Lee J, Roh S, Shin J, Sohn K (2019b) Image-based learning to measure the space mean speed on a stretch of road without the need to tag images with labels. Sensors 19:1227

Li Y, Møgelmose A, Trivedi MM (2016) Pushing the "Speed Limit": high-accuracy US traffic sign recognition with convolutional neural networks. IEEE Trans Intell Veh 1:167–176

Li Y, Yu R, Shahabi C, Liu Y (2018) Diffusion convolutional recurrent neural network: data-driven traffic forecasting. In: International conference on learning representations

Li L, Qin L, Qu X, Zhang J, Wang Y, Ran B (2019a) Day-ahead traffic flow forecasting based on a deep belief network optimized by the multi-objective particle swarm algorithm. Knowl Based Syst 172:1–14

Li Z, Yang Q, Chen S, Zhou W, Chen L, Song L (2019b) A fuzzy recurrent neural network for driver fatigue detection based on steering-wheel angle sensor data. Int J Distrib Sens Netw 15:1550147719872452

Li L, Qu X, Zhang J, Wang Y, Ran B (2019c) Traffic speed prediction for intelligent transportation system based on a deep feature fusion model. J Intell Transp Syst 23(6):605–616

Lin Y, Dai X, Li L, Wang F-Y (2018) Pattern sensitive prediction of traffic flow based on generative adversarial framework. IEEE Trans Intell Transp Syst 20:2395–2400

Liu L, Chen R-C (2017a) A novel passenger flow prediction model using deep learning methods. Transp Res Part C Emerg Technol 84:74–91. https://doi.org/10.1016/j.trc.2017.08.001

Liu L, Chen R-C (2017b) A MRT daily passenger flow prediction model with different combinations of influential factors. In: Advanced information networking and applications workshops (WAINA), 2017 31st international conference on. IEEE, pp 601–605

Liu Q, Wang B, Zhu Y (2018) Short-term traffic speed forecasting based on attention convolutional neural network for arterials. Comput Civ Infrastruct Eng 33:999–1016

Liu Y, Liu Z, Jia R (2019) DeepPF: a deep learning based architecture for metro passenger flow prediction. Transp Res Part C Emerg Technol 101:18–34

Luo P, Tian Y, Wang X, Tang X (2014) Switchable deep network for pedestrian detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 899–906

Lv Y, Duan Y, Kang W, Li Z, Wang F-Y (2015) Traffic flow prediction with big data: a deep learning approach. IEEE Trans Intell Transp Syst 16:865–873

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

Ma X, Yu H, Wang Y, Wang Y (2015b) Large-scale transportation network congestion evolution prediction using deep learning theory. PLoS ONE 10:e0119044

Ma X, Dai Z, He Z, Ma J, Wang Y, Wang Y (2017) Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction. Sensors. https://doi.org/10.3390/s17040818

Mackenzie J, Roddick JF, Zito R (2018) An evaluation of HTM and LSTM for short-term arterial traffic flow prediction. IEEE Trans Intell Transp Syst 20:1847–1857

Ministry of Internal Affairs and Communications (2019) AI development guidelines

Mou L, Zhao P, Xie H, Chen Y (2019) T-LSTM: a long short-term memory neural network enhanced by temporal information for traffic flow prediction. IEEE Access 7:98053–98060

Moussavi-Khalkhali A, Jamshidi M (2016) Constructing a deep regression model utilizing cascaded sparse autoencoders and stochastic gradient descent. In: 2016 15th IEEE international conference on machine learning and applications (ICMLA). IEEE, pp 559–564

Nguyen H, Kieu L-M, Wen T, Cai C (2018) Deep learning methods in transportation domain: a review. IET Intell Transp Syst 12:998–1004

Ouyang W, Wang X (2013) Joint deep learning for pedestrian detection. In: Proceedings of the IEEE international conference on computer vision, pp 2056–2063

Pamuła T (2018) Impact of data loss for prediction of traffic flow on an urban road using neural networks. IEEE Trans Intell Transp Syst 20(3):1000–1009

Petersen NC, Rodrigues F, Pereira FC (2019) Multi-output bus travel time prediction with convolutional LSTM neural network. Expert Syst Appl 120:426–435

Ping P, Sheng Y, Qin W, Miyajima C, Takeda K (2018) Modeling driver risk perception on city roads using deep learning. IEEE Access 6:68850–68866. https://doi.org/10.1109/ACCESS.2018.2879887

Polson NG, Sokolov VO (2017) Deep learning for short-term traffic flow prediction. Transp Res Part C Emerg Technol 79:1–17. https://doi.org/10.1016/j.trc.2017.02.024

Qian R, Zhang B, Yue Y, Wang Z, Coenen F (2015) Robust chinese traffic sign detection and recognition with deep convolutional neural network. In: 2015 11th international conference on natural computation (ICNC). IEEE, pp 791–796

Qin Y, Luo H, Zhao F, Zhao Z, Jiang M (2018) A traffic pattern detection algorithm based on multimodal sensing. Int J Distrib Sens Netw 14:1550147718807832

Qu L, Li W, Li W, Ma D, Wang Y (2019) Daily long-term traffic flow forecasting based on a deep neural network. Expert Syst Appl 121:304–312

Ran X, Shan Z, Fang Y, Lin C (2019a) A convolution component-based method with attention mechanism for travel-time prediction. Sensors 19:2063

Ran X, Shan Z, Shi Y, Lin C (2019b) Short-term travel time prediction: a spatiotemporal deep learning approach. Int J Inf Technol Decis Mak 18:1087–1111

Ren H, Song Y, Liu J, Hu Y, Lei J (2017) A deep learning approach to the prediction of short-term traffic accident risk. arXiv1710.09543

Ren Y, Cheng T, Zhang Y (2019) Deep spatio-temporal residual neural networks for road-network-based data modeling. Int J Geogr Inf Sci 33(9):1894–1912

Sameen M, Pradhan B (2017) Severity prediction of traffic accidents with recurrent neural networks. Appl Sci 7:476

Shafique MA, Hato E (2015) Use of acceleration data for transportation mode prediction. Transportation (Amst) 42:163–188

Sifringer B, Lurkin V, Alahi A (2018) Enhancing discrete choice models with neural networks. In: HEART 2018—7th symposium of the european association for research in transportation conference

Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv1409.1556

Singh D, Mohan CK (2018) Deep spatio-temporal representation for detection of road accidents using stacked autoencoder. IEEE Trans Intell Transp Syst 20(3):879–887

Siripanpornchana C, Panichpapiboon S, Chaovalit P (2016) Travel-time prediction with deep learning. In: 2016 IEEE region 10 conference (TENCON). IEEE, pp 1859–1862

Soua R, Koesdwiady A, Karray F (2016) Big-data-generated traffic flow prediction using deep learning and dempster-shafer theory. In: Neural networks (IJCNN), 2016 international joint conference on. IEEE, pp 3195–3202

Sun F, Dubey A, White J (2017) DxNAT—deep neural networks for explaining non-recurring traffic congestion. In: 2017 IEEE international conference on big data (big data). IEEE, pp 2141–2150

Sun S, Chen J, Sun J (2019) Traffic congestion prediction based on GPS trajectory data. Int J Distrib Sens Netw 15:1550147719847440

Tan H, Xuan X, Wu Y, Zhong Z, Ran B (2016) A comparison of traffic flow prediction methods based on DBN. CICTP 2016:273–283

Tang K, Chen S, Khattak AJ, Pan Y (2019a) Deep architecture for citywide travel time estimation incorporating contextual information. J Intell Transp Syst. https://doi.org/10.1080/15472450.2019.1617141

Tang Q, Yang M, Yang Y (2019b) ST-LSTM: a deep learning approach combined spatio-temporal features for short-term forecast in rail transit. J Adv Transp 2019:8392592. https://doi.org/10.1155/2019/8392592

Tian Y, Pan L (2015) Predicting short-term traffic flow by long short-term memory recurrent neural network. In: 2015 IEEE international conference on smart city/SocialCom/SustainCom (SmartCity). IEEE, pp 153–158

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

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

Tian D, Zhang C, Duan X, Wang X (2019) An automatic car accident detection method based on cooperative vehicle infrastructure systems. IEEE Access 7:127453–127463

Torres R, Ohashi O, Pessin G (2019) A machine-learning approach to distinguish passengers and drivers reading while driving. Sensors 19:3174

Tran D, Do HM, Sheng W, Bai H, Chowdhary G (2018) Real-time detection of distracted driving based on deep learning. IET Intell Transp Syst 12:1210–1219

Tu W, Xiao F, Fu L, Pan G (2017) A deep learning model for traffic flow state classification based on smart phone sensor data. arXiv1709.08802

Wang J, Shi Q (2013) Short-term traffic speed forecasting hybrid model based on chaos–wavelet analysis-support vector machine theory. Transp Res Part C Emerg Technol 27:219–232

Wang J, Gu Q, Wu J, Liu G, Xiong Z (2016) Traffic speed prediction and congestion source exploration: a deep learning method. In: 2016 IEEE 16th international conference on data mining (ICDM), pp 499–508. https://doi.org/10.1109/ICDM.2016.0061

Wang J, Hu F, Li L (2017) Deep bi-directional long short-term memory model for short-term traffic flow prediction. In: International conference on neural information processing. Springer, pp 306–316

Wang Y, Geng S, Gao H (2018a) A proactive decision support method based on deep reinforcement learning and state partition. Knowl Based Syst 143:248–258. https://doi.org/10.1016/j.knosys.2017.11.005

Wang P, Hao W, Sun Z, Wang S, Tan E, Li L, Jin Y (2018b) Regional detection of traffic congestion using in a large-scale surveillance system via deep residual TrafficNet. IEEE Access 6:68910–68919

Wang Y, Zhang D, Liu Y, Dai B, Lee LH (2019) Enhancing transportation systems via deep learning: a survey. Transp Res Part C Emerg Technol 99:144–163. https://doi.org/10.1016/j.trc.2018.12.004

Wee BV, Banister D (2016) How to write a literature review paper? Transp Rev 36:278–288

Willis C, Harborne D, Tomsett R, Alzantot M (2017) A deep convolutional network for traffic congestion classification. In: Proceedings of NATO IST-158/RSM-010 specialists' meeting on content based real-time analytics of multi-media streams, pp 1–11

Wu Y, Tan H (2016) Short-term traffic flow forecasting with spatial-temporal correlation in a hybrid deep learning framework. arXiv1612.01022

Wu C-H, Ho J-M, Lee D-T (2004) Travel-time prediction with support vector regression. IEEE Trans Intell Transp Syst 5:276–281

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

Xiangxue W, Lunhui X, Kaixun C (2019) Data-driven short-term forecasting for urban road network traffic based on data processing and LSTM-RNN. Arab J Sci Eng 44:3043–3060

Xie D-F, Fang Z-Z, Jia B, He Z (2019) A data-driven lane-changing model based on deep learning. Transp Res Part C Emerg Technol 106:41–60

Xing Y, Lv C, Wang H, Cao D, Velenis E, Wang FY (2019) Driver activity recognition for intelligent vehicles: a deep learning approach. IEEE Trans Veh Technol 68(6):5379–5390

Xu C, Ji J, Liu P (2018a) The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets. Transp Res Part C Emerg Technol 95:47–60. https://doi.org/10.1016/j.trc.2018.07.013

Xu C, Ji J, Liu P (2018b) The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets. Transp Res Part C Emerg Technol 95:47–60. https://doi.org/10.1016/j.trc.2018.07.013

Xu T, Li X, Claramunt C (2018c) Trip-oriented travel time prediction (TOTTP) with historical vehicle trajectories. Front Earth Sci 12:253–263

Xu J, Rahmatizadeh R, Bölöni L, Turgut D (2018d) Real-time prediction of taxi demand using recurrent neural networks. IEEE Trans Intell Transp Syst 19:2572–2581

Yang H-F, Chen Y-PP (2019) Hybrid deep learning and empirical mode decomposition model for time series applications. Expert Syst Appl 120:128–138

Yang H-F, Dillon TS, Chen Y-PP (2017) Optimized structure of the traffic flow forecasting model with a deep learning approach. IEEE Trans Neural Netw Learn Syst 28:2371–2381

Yang S, Ma W, Pi X, Qian S (2019) A deep learning approach to real-time parking occupancy prediction in transportation networks incorporating multiple spatio-temporal data sources. Transp Res Part C Emerg Technol 107:248–265

Yang G, Wang Y, Yu H, Ren Y, Xie J (2018) Short-term traffic state prediction based on the spatiotemporal features of critical road sections. Sensors. https://doi.org/10.3390/s18072287

Yao H, Wu F, Ke J, Tang X, Jia Y, Lu S, Gong P, Ye J (2018) Deep multi-view spatial-temporal network for taxi demand prediction. arXiv1802.08714

Yi, H, Jung H, Bae S (2017) Deep neural networks for traffic flow prediction. In: 2017 IEEE international conference on big data and smart computing (BigComp). IEEE, pp 328–331

Yogameena B, Menaka K, Perumaal SS (2019) Deep learning-based helmet wear analysis of a motorcycle rider for intelligent surveillance system. IET Intell Transp Syst 13(7):1190–1198

Yu G, Liu J (2019) A hybrid prediction approach for road tunnel traffic based on spatial-temporary data fusion. Appl Intell 49:1421–1436

Yu H, Wu Z, Wang S, Wang Y, Ma X (2017a) Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks. Sensors 17:1501

Yu B, Yin H, Zhu Z (2017b) Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting. arXiv1709.04875

Yu R, Li Y, Shahabi C, Demiryurek U, Liu Y (2017c) Deep learning: a generic approach for extreme condition traffic forecasting. In: Proceedings of the 2017 SIAM international conference on data mining. SIAM, pp 777–785

Zang D, Ling J, Cheng J, Tang K, Li X (2017a) Using convolutional neural network with asymmetrical kernels to predict speed of elevated highway. In: International conference on intelligence science. Springer, pp 212–221

Zang D, Wang D, Cheng J, Tang K, Li X (2017b) Traffic parameters prediction using a three-channel convolutional neural network. In: International conference on intelligence science. Springer, pp 363–371

Zang D, Fang Y, Wei Z, Tang K, Cheng J (2019) Traffic flow data prediction using residual deconvolution based deep generative network. IEEE Access

Zhang Q, Zhu S-C (2018) Visual interpretability for deep learning: a survey. Front Inf Technol Electron Eng 19:27–39

Zhang Z, He Q, Gao J, Ni M (2018) A deep learning approach for detecting traffic accidents from social media data. Transp Res Part C Emerg Technol 86:580–596

Zhang Z, Li M, Lin X, Wang Y, He F (2019a) Multistep speed prediction on traffic networks: a deep learning approach

considering spatio-temporal dependencies. Transp Res Part C Emerg Technol 105:297–322

Zhang X, Sun J, Qi X, Sun J (2019b) Simultaneous modeling of car-following and lane-changing behaviors using deep learning. Transp Res Part C Emerg Technol 104:287–304

Zhang J, Wu Z, Li F, Xie C, Ren T, Chen J, Liu L (2019c) A deep learning framework for driving behavior identification on in-vehicle CAN-BUS sensor data. Sensors 19:1356

Zhang Y, Cheng T, Ren Y (2019d) A graph deep learning method for short-term traffic forecasting on large road networks. Comput-Aided Civ Infrastruct Eng 34(10):877–896

Zhao D, Dai Y, Zhang Z (2011) Computational intelligence in urban traffic signal control: a survey. IEEE Trans Syst Man Cybern Part C (Appl Rev) 42(4):485–494

Zhao Z, Chen W, Wu X, Chen PCY, Liu J (2017) LSTM network: a deep learning approach for short-term traffic forecast. IET Intell Transp Syst 11:68–75

Zhao J, Gao Y, Qu Y, Yin H, Liu Y, Sun H (2018) Travel time prediction: based on gated recurrent unit method and data fusion. IEEE Access 6:70463–70472

Zhao J, Gao Y, Bai Z, Wang H, Lu S (2019a) Traffic speed prediction under non-recurrent congestion: based on LSTM method and BeiDou navigation satellite system data. IEEE Intell Transp Syst Mag 11:70–81

Zhao W, Gao Y, Ji T, Wan X, Ye F, Bai G (2019b) Deep temporal convolutional networks for short-term traffic flow forecasting. IEEE Access 7:114496–114507. https://doi.org/10.1109/ACCESS.2019.2935504

Zhao J, Gao Y, Yang Z, Li J, Feng Y, Qin Z, Bai Z (2019c) Truck traffic speed prediction under non-recurrent congestion: based on optimized deep learning algorithms and GPS data. IEEE Access 7:9116–9127

Zhao H, Mao T, Duan J, Wang Y, Zhu H (2019d) FMCNN: a factorization machine combined neural network for driving

safety prediction in vehicular communication. IEEE Access 7:11698–11706

Zhao Lu, Zhou Y, Lu H, Fujita H (2019e) Parallel computing method of deep belief networks and its application to traffic flow prediction. Knowl Based Syst 163:972–987. https://doi.org/10.1016/j.knosys.2018.10.025

Zhao Hong, Hou C, Alrobassy H, Zeng X (2019f) Recognition of transportation state by smartphone sensors using deep Bi-LSTM neural network. J Comput Netw Commun 2019:4967261. https://doi.org/10.1155/2019/4967261

Zheng M, Li T, Zhu R, Chen J, Ma Z, Tang M, Cui Z, Wang Z (2019) Traffic accident's severity prediction: a deep-learning approach-based CNN network. IEEE Access 7:39897–39910

Zhou T, Han G, Xu X, Han C, Huang Y, Qin J (2019) A learning-based multimodel integrated framework for dynamic traffic flow forecasting. Neural Process Lett 49:407–430

Zhu L, Laptev N (2017) Deep and confident prediction for time series at uber. In: Data mining workshops (ICDMW), 2017 IEEE international conference on. IEEE, pp 103–110

Zhu X, Li J, Liu Z, Yang F (2017) Location deployment of depots and resource relocation for connected car-sharing systems through mobile edge computing. Int J Distrib Sens Netw 13:1550147717711621

Zhu J, Huang C, Yang M, Fung GPC (2019) Context-based prediction for road traffic state using trajectory pattern mining and recurrent convolutional neural networks. Inf Sci (N Y) 473:190–201