A comprehensive study of speed prediction in transportation system: From vehicle to traffic

Zewei Zhou,1 Ziru Yang,1 Yuanjian Zhang,2 Yanjun Huang,1,* Hong Chen,3 and Zhuoping Yu1

SUMMARY
In the intelligent transportation system (ITS), speed prediction plays a significant role in supporting vehicle routing and traffic guidance. Recently, a considerable amount of research has been devoted to a single-level (e.g., traffic or vehicle) prediction. However, a systematic review of speed prediction in and between different levels is still missing. In this article, existing research is comprehensively analyzed and divided into three levels, i.e. macro traffic, micro vehicles, and meso lane. In addition, this article summarizes the influencing factors and reviews the prediction methods based on how those methods utilize the available information to meet the challenges of the prediction at different levels. This is followed by a summary of evaluation metrics, public datasets, and open-source codes. Finally, future directions in this field are discussed to inspire and guide readers. This article aims to draw a complete picture of speed prediction and promote the development of ITS.

INTRODUCTION
The transportation system is the blood vessel of cities. However, with the rapid gathering of population toward cities, the problems of urban traffic, such as congestion, pollution, and accidents, have significantly affected the traffic efficiency and city development. To address these issues, intelligent transportation system (ITS) has attracted extensive interests (Yuan and Li, 2021), which provides efficient traffic service and management. Different from constructing new infrastructures (e.g., roads and bridges), limited by space and high cost (Lana et al., 2018), ITS can collect and process diverse data through intelligent infrastructure and advanced algorithms to improve traffic efficiency. The prediction of the traffic states online or offline is a fundamental part of ITS (Nagy and Simon, 2018), which enables the current traffic service and management far sight. Among the traffic states, speed is the basic property of dynamic traffic, which reflects vehicle motion and traffic efficiency (Qu et al., 2021). Furthermore, route planning and traffic intervention in advance based on speed prediction is an effective way to improve traffic efficiency, reduce energy consumption, and improve participants’ experience.

Speed prediction aims to estimate the speed of traffic participants in a future period based on current and historical traffic states. Meanwhile, prediction can be classified into three categories according to the different scales, i.e., traffic speed prediction (macro), vehicle speed prediction (micro), lane-level speed prediction (meso). Traffic flow consisting of multiple vehicles as the target of traffic-speed prediction describes the dynamic characteristics of traffic in a macroscopic view. Thus, the time horizon of this prediction can range from minutes to hours, even days (Zang et al., 2019), as shown in Table 1. The accurate short-term prediction of traffic speed in minutes can support traffic management, such as the optimization of signal timing and traffic resources allocation (Wang et al., 2016). Moreover, traffic participants can utilize the traffic trend to plan their travel (Park et al., 2014). On the other hand, the traffic patterns captured by long-term prediction in hours help to understand the traffic and support transportation planning (Nagy and Simon, 2018). For example, the prediction methods can capture the main factors causing congestion in a specific scenario and support the planning of new roads to relieve congestion. However, vehicle speed prediction focuses on the identification of the future micro speed patterns. The future information of vehicle speed is useful for trajectory planning and eco-driving (Huang et al., 2018; Ye et al., 2019). Especially for the hybrid electric vehicles, the speed prediction is a key part of energy management strategy (EMS) to improve the powertrain efficiency (Sun et al., 2015a; Liu et al., 2021). Collision risk can be reduced by a reliable speed prediction as well (Liu et al., 2019). The traffic flow of lanes at mesoscale, as the target of lane-level prediction, describes the dynamic traffic in more detail than macro traffic speed and captures the interaction between individual vehicles.
between lanes, such as lane change. Therefore, it is able to provide the speed information at lane level, which is essential for lane-based traffic applications such as high-precision navigation and intelligent driving (Gu et al., 2019b; Shen et al., 2021). Table 1 shows the detailed comparison of the speed prediction at different levels in terms of the prediction target, speed characteristics, time horizon, and application.

Because of the influence of various contextual factors, e.g., weather and accidents, the traffic is a chaotic dynamic system in essence, which brings complex nonlinearity and uncertainty to speed variables. Therefore, how to discover the hidden patterns in traffic is the most critical problem for speed prediction. To address this, plenty of prediction methods have been proposed, and the first category belongs to the theoretical model-based methods. Such deductive methods (Wang and Shi, 2013) are easy to realize but inevitably suffer from inaccuracy because of strong artificial assumptions (Shen et al., 2018a). In addition, those methods are powerless to describe various factors and model their uncertainty. Unlike the model-based methods, the data-driven ones inductively learn the law from available data and capture the traffic pattern without explicit models. Such methods can model various factors gracefully and obtain high-quality results. However, the interpretability and adaptability of such methods are still a puzzle. With the rapid development of data acquisition and computing power, the research priority of speed prediction gradually transforms from analytical methods to data-driven methods (Vlahogianni et al., 2014). Meanwhile, the data-driven methods have evolved from statistical methods to traditional machine learning and recently to deep-learning methods. Because of its powerful capacities to fit non-linear functions and analyze big

| Target | Average Speed of Multiple Vehicles in Road & Network | Average Speed of Multiple Vehicles in Road Lane | Single Vehicle Speed |
|--------|-----------------------------------------------------|------------------------------------------------|----------------------|
| Characteristic | External Factors Influence | Interactions between Lanes | Driver and Vehicle Influence |
| Dynamic Spatio-Temporal Dependency | Lane-level Spatio-Temporal Dependency | Dynamic Temporal Dependency |
| Time Horizon | 05 min | 01-30 s | 05 s |
| | 15 min | 02-10 min | 15 s |
| | 30 min | 10-20 min | 30 s |
| | 60 min | 30 min | 50 s |
| Application | Traffic Management | High-precision Traffic Management | Eco-Driving(Energy and Thermal Management) |
| (Traffic Resources Allocation and Signal Timing Optimization) | | | |
| | Travel Planning (Efficient and Comfortable) | Lane-level Travel Planning | Collision Risk Estimation |
| Transportation Planning | | | |
| | Nagy and Simon, 2018 | Ke et al., 2018a, 2020b | Vogel, 2003 |
| | Liao et al., 2018 | | Polychronopoulos et al., 2007 |
| | | | Barrios and Motai, 2011 |
data, deep-learning methods are considered as the most promising approach (Yin et al., 2021a). In addition, because the characteristics of speed predictions at different levels are not the same, specific methods should meet the different demands of speed prediction in applications.

**Difference from existing surveys on speed prediction**

Regarding traffic-speed prediction, reference (Vlahogianni et al., 2014) summarized ten challenges based on the research from 2004 to 2013. However, it could not cover the popular methods based on deep learning. Reference (Yuan and Li, 2021) provided a survey from the data layer to application layer of traffic prediction, and the traffic data format was discussed in detail in (Nagy and Simon, 2018). But the prediction methods are not the emphasis in their discussion. Thus, study (Migliani and Kumar, 2019) presented a survey of deep-learning methods used for traffic prediction in autonomous vehicles, and study (Tedjopurnomo et al., 2020) focused on three common neural network methods. However, those researches paid insufficient attention to recent advances, e.g., attention and graph-based learning models. The deep-learning methods were categorized into five generations to describe the research trend in (Lee et al., 2021), and reference (Yi et al., 2020) provided a survey specifically on graph-based deep learning. Meanwhile, research (Yin et al., 2021a) conducted experiments to compare different deep learning methods, and research (López Manibardo et al., 2021) discussed the pros and cons of deep-learning methods in detail. Most of these works focus on traffic prediction (e.g., flow, speed, and demand prediction), but the detailed review of speed prediction is still absent.

With respect to vehicle speed prediction, parametric and non-parametric methods were compared over 1–10 s horizon in (Lefèvre et al., 2014), which indicated that non-parametric methods are suitable for the long-term prediction. Reference (Liu et al., 2019) classified the prediction methods of vehicle speed into two categories: deterministic and stochastic prediction, and reference (Zhou et al., 2019c) provided an overview of driving prediction (e.g., speed, acceleration) and its application. Meanwhile, study (Huang et al., 2017) presented a review of model predictive control based strategies in EMS, and simply reviewed the prediction methods of vehicle-speed. The existing works about vehicle speed prediction focus on the application of EMS, but the systematic discussion of the prediction methods is still missing. Furthermore, there are fewer studies on lane-level speed prediction since it has been enjoying attention in recent years. In conclusion, most existing literature focuses on the speed prediction only at a single level. The community lacks a systematic and comprehensive survey to explore the differences and similarities of speed prediction at different levels. Our research aims to fill this gap and inspire future researchers.

**The contributions of this article**

- This article systematically summarizes the speed prediction in the transportation system into three categories according to the scale: Traffic (macro), vehicle (micro), and lane-level (meso). The similarities and differences of speed prediction at different levels are explored to promote a comprehensive understanding.
- Various influencing factors of speed are summarized and a review of different speed predictions is provided based on the type and amount of information utilized. Meanwhile, this article describes the theory, characteristics, and variants of each method to inspire further research.
- The existing evaluation metrics for prediction models are collected, and the public traffic datasets and open-source models are also organized to facilitate future work’s development and experimentation.
- This article concludes with a discussion of the challenges and future directions for the speed prediction problem in transportation.

**Organization of this article**

Definitions and preliminaries section presents the definition and characteristic of speed prediction at different levels. The influencing factors of speed are also demonstrated. Besides, Prediction methods of traffic speed section Prediction methods of vehicle speed section, and Lane-level speed prediction section provide an overview of prediction methods at different levels in terms of model-based and data-driven methods. Evaluation section provides evaluation metrics, public datasets, and codes in the literature.
and discusses the existing challenges and future directions. Finally, Conclusion section draws a conclusion of this article.

DEFINITIONS AND PRELIMINARIES

The speed prediction aims to obtain the future speed as accurately as possible over a specific time horizon. The essence of speed prediction is the identification of future traffic patterns, which describes the traffic operation mode and the relationship between the future speed and historical traffic states. Two important factors determine the prediction performance: the model and available information. The former takes charge of extracting patterns, which will be described detailed in Section 3-5. The latter, as the input of prediction model, is the foundation of prediction and can significantly affect the prediction accuracy. In addition, the features of information directly determine the information quality, such as data type, sources (i.e., fixed point data, vehicle trajectory data), format (i.e., scalar, vector, matrix), and processing (i.e., map matching, data cleaning). This article focuses on the type and amount of information utilized in speed prediction. This section begins with the definition of speed prediction in the transportation system, which explores the similarities and differences of speed prediction at different levels. Then, the influencing factors of speed are reviewed, and the main factors of the speed at different levels are analyzed.

Definition of speed prediction at different levels

On the one hand, traffic flow exhibits periodicity and consistency. On the other hand, it is volatile and chaotic with the influence of contextual factors. Whether speed in the transportation system can be predicted and whether a pattern worth being captured exists are the basis of predictability. Predictability is defined as the possibility of achieving the desired prediction accuracy over a specific horizon (Yue et al., 2007). Overall, the speed at the proposed three levels is predictable, which is confirmed by a large amount of work on speed prediction (Wang and Shi, 2013; Lefevre et al., 2014; Gu et al., 2019b).

Given a prediction problem over a certain horizon, the ground truth can be divided into deterministic and uncertain part (Yue et al., 2007) as shown in Figure 1. In this article, the deterministic part refers to the trend and range of speed, which is determined by the available information. For example, the speed of an accelerating vehicle is likely to increase at the next moment and impossible to slow down instantly. Meanwhile, the uncertainty caused by contextual factors reflects the stochastic and chaotic characteristics of the traffic system. Parts of uncertainty are predictable, but the rest is unforeseeable depending on how well it is understood (Yue et al., 2007). Thus, the theoretically predictable value consists of the deterministic and the predictable portion in the uncertainty part (Nagy and Simon, 2018). The information basis determines the theoretically predictable value and determines the ceiling of prediction accuracy. In practice, the model used for prediction inevitably contains an error, as the three parts of error shows in Figure 1. Therefore, considering the various contextual factors it is essential to increase the theoretically predictable value, and the improvement of performance is the key to decreasing the error. In the context of predictability analysis, each speed prediction is defined as follows.

Traffic speed prediction

Traffic speed prediction is represented by the average speed of multiple vehicles passing through a specific road section over a period. The physical structure of road networks brings spatial and temporal
dependency to traffic speed, which is key to improving prediction accuracy (Ma et al., 2017). However, how to model the dependency is still a challenge (Zheng et al., 2020).

**Spatial dependency.** Traffic flows follow the road network, where the speeds of spatial points in the vicinity are correlated. Different spatial points have different effects on the prediction result and the spatial dependency at different time is different. The spatial dependency can be divided into local and global ones (Yin et al., 2021b). The former focuses on the local neighbor roads and the other concerns the connectivity of the entire network. In addition, a region in the road network is usually space-dependent with another through various non-Euclidean relation such as spatial adjacency, Point of Interest (POI) (Ge et al., 2019), and semantic information (Liao et al., 2018). As shown in Figure 2A, the three marked regions are connected and spatial adjacent to each other through the yellow road network. The POI attributes of the three marked regions also affect the corresponding traffic pattern.

**Temporal dependency.** The current speed at a specific point is correlated with the historical speed. Temporal dependency is complex non-linear and differs from one point to another. In addition, different time slices of speed bring different effects on the current speed (Wu et al., 2018). Moreover, traffic speeds are cyclical in the temporal dimension; for example, traffic speed is similar on weekdays and different from weekends. An important representation of this similarity on the weekday is the morning and evening rush, which is denoted by the red and blue arrows as shown in Figure 2A.

**Vehicle speed prediction**

Vehicle speed prediction focuses on a single vehicle as shown in Figure 2B. No direct spatial dependency is exhibited in vehicle speeds. Besides, the obvious uncertainty of vehicle behavior brings a more complex temporal dependency than traffic speed. Moreover, the prediction is also characterized by a short prediction horizon and strict real-time requirements, as vehicle speed prediction is mainly used for energy management (Zhang, 2019) and thermal management (Wang and Infield, 2018) of eco-driving.

**Lane-level speed prediction**

Lane-level speed prediction aims to obtain the average speed of vehicles passing through a certain lane cross-section within a certain period. Figure 3 shows a brief description of the lane-level scenario. Unlike the macroscopic or microscopic speed prediction, the granularity of lane-level speed prediction is meso-scale (Ke et al., 2020a). The characteristics of this prediction lie between traffic and vehicle-speed prediction. Moreover, the modeling of interaction and spatio-temporal dependency at lane level are the challenges of this prediction.

**Influencing factors on speed prediction**

In this article, factors influencing speed are summarized according to two principles: internal/external and static/dynamic. The former is classified by vehicle perspective, which represents the internal factors of the vehicle and the external factors of the environment, as shown in Figures 2B and 3. In the latter classification, static factors are constant and can be accessed online through offline storage. On the other hand, dynamic
factors vary over time such that it is more difficult to model them than static ones. Thus, historical and current information of dynamic factors usually should be considered together in prediction. Finally, this section analyzes the main influencing factors of different predictions.

**Internal factors**

**Driving behavior (dynamic).** The human driver directly operates the vehicle, and his/her driving behavior significantly affects vehicle speed. All the other factors affect the speed indirectly through driving behavior. However, the perception and decision of human driver are extremely complex (Lefevre et al., 2014; Lian et al., 2017). Drivers may be on the phone, talking, and becoming fatigued while driving. Furthermore, in the same scenario, the responses of different drivers are not similar, and even the same driver possibly makes different decisions (Yeon et al., 2019; Moser et al., 2015).

**Vehicle information (static).** Vehicle information refers to the static factors about vehicle, including physical information (e.g., vehicle type, vehicle weight, passenger weight, and distribution), powertrain (e.g., power type, transmission type), power performance (e.g., maximum acceleration/deceleration, maximum speed, maximum climb), handling stability (e.g., minimum turning radius), etc. In addition, vehicle functions (e.g., ADAS) also belong to such information, which controls the vehicle speed according to specific rules.

**Vehicle state (dynamic).** The above vehicle information constrains the change of vehicle speed indirectly, and the dynamic vehicle states directly affect speed. These factors describe the running state of vehicle, including vehicle speed, acceleration, available fuel or power (Sun et al., 2015a), battery temperature (Amini et al., 2020), transmission state (Li et al., 2018a), etc.

**External factors**

**Traffic flow state (dynamic).** From a microscopic perspective, the traffic flow state refers to the moving state of surrounding vehicles, and the front vehicle states (i.e., relative speed and distance) heavily impact ego-vehicle speed (Yeon et al., 2019; Suh et al., 2020). From a macroscopic perspective, it refers to the traffic state, e.g., flow, speed, and occupancy (Liu et al., 2018a; Zhou et al., 2019a; Yan et al., 2018). For example, the traffic congestion can directly affect the speed at different levels, and the macroscopic shock waves are an influencing factor. In addition, the traffic flow state shows a strong spatio-temporal dependency, and the historical traffic flow state should be considered.

**Weather conditions (dynamic).** The bad weather conditions (e.g., fog, rain, and snow) significantly affect the visibility and road friction, such that change the driving behaviors (Ahmed and Ghasemzadeh, 2018) and vehicle stability (Yang et al., 2020). For example, the low visibility caused by heavy fog or rain heavily impact the perception and control of vehicles (Ahmed and Ghasemzadeh, 2018; Fridman et al., 2019). Therefore, these factors may cause traffic congestion or accidents, which seriously affect the speed at different level.
Road and traffic rules (static). Roads include urban roads, highways, rural roads and some functional roads such as ramps (Li et al., 2019b), intersections, transitional sections, sharp turns (Liu et al., 2018b; Qian et al., 2020; Yeon et al., 2019). The road properties (i.e., slope (Gu et al., 2019a), super elevation, curvature and roughness), environmental attributes (e.g., cluster, schools, hospitals) and traffic rules (e.g., speed hump, speed limits, lane properties) have an impact on speed (Shin and Sunwoo, 2019).

Traffic signals and events (dynamic). Traffic events include traffic accidents (Xie et al., 2019a; Qian et al., 2020), traffic control measures, social events (e.g., sporting events, examinations, performances) (Polson and Sokolov, 2017), etc. The traffic accidents may lead to congestion, and the social events affect the speed by changing the traffic demand. In addition, traffic signals are the essential to keeping traffic in order on urban roads, and it severely constrains the speed in the transportation system (Zhou et al., 2019a). Furthermore, the lighting timing is dynamic thanks to real-time traffic management.

Factor analysis at different levels. All the above influencing factors impact the speed at different levels, but the main factors at each level are not the same. The micro speed refers to the vehicle speed such that the internal factors about the single-vehicle states are the main factors of vehicle speed in short-term prediction, especially the vehicle behavior. Besides, with the increase of prediction horizon, the impact of internal factors could decrease and the impact of external factors could increase thanks to their cumulative effect. With regard to the traffic speed, the effect of single-vehicle uncertainty on macro speed is attenuated by the statistical averaging over multiple vehicles. Therefore, compared to vehicle speed, the traffic speed varies steadily and the impact of internal factors on it is less. Meanwhile, the external factors become the main factors of traffic speed thanks to its longer prediction horizon than that of vehicle speed. As for lane-level speed, the impact of external and internal factors is between the macro and micro speed because of its mesoscopic perspective.

PREDICTION METHODS OF TRAFFIC SPEED

The spatio-temporal dependency is the main characteristic of traffic speed, such that it is the focus of prediction methods modeling. Meanwhile, integrating various influencing factors into model is the trend of prediction. Table 2 shows the summary and taxonomy of the method used for traffic speed prediction. We review the methods in terms of model-based, classical data-driven, and deep-learning methods. Meanwhile, the information utilized and how to model them is the concern of this section.

Model-based methods

The model-based methods designed for traffic-speed prediction can be categorized according to the granularity as macroscopic, microscopic, and mesoscopic. The three levels are similar to the classification used to discuss the speed prediction problem in this article, but the specific definitions are different.

Macroscopic methods utilize macroscopic traffic properties (i.e. average speed, average traffic flow, and density) to describe traffic and analyze the traffic behavior through traffic flow theory. A classical macroscopic method is kinematic wave (Newell, 1993) which can reproduce the propagation of traffic waves. However, junctions are the bottlenecks of the kinematic wave method where multiple traffic waves meet. Reference (Jin, 2010) used the concepts of demand and supply to formulate the merging in kinematic wave and obtained analytical solutions. Meanwhile, the merging process was regarded as fair queuing based on their capacity in (Ni and Leonard, 2005). In addition, the choice of coordinate system in such methods directly affects prediction accuracy. Reference (Van Wageningen-Kessels et al., 2010) proposed a Lagrangian coordinate system that moves with the vehicle, resulting in a more accurate result than the traditional coordinate system.

Microscopic model-based methods aim to simulate traffic with the detailed responses of traffic participants via modeling their behavior. The intention of participant and interaction between each determine its behavior such that the Cellular Automata with the advantage of modeling the interaction is popular (Nagel and Schreckenberg, 1992; Maiervoet and DeMoor, 2005). Reference (Korcik et al., 2011) eliminated the unwanted properties of Cellular Automata and applied the model in the speed prediction of large-scale networks. Considering the influence of special vehicles, reference (Zhao et al., 2020a) introduced some special rules for Cellular Automata (Zhao et al., 2020a), such as the low reaction time of connected-automated vehicle and regional avoidance of emergency vehicles.
| Model type | Modeling traffic spatial dependency | Modeling traffic temporal dependency | Modeling external factors | Reference |
|------------|-----------------------------------|-----------------------------------|--------------------------|-----------|
| Statistical Methods | ARIMA, SARIMA | – | Williams and Hoel, 2003; Lippi et al., 2013 |
| | VARIMA, STARIMA | – | Chandra and Al-Deek, 2009; Min and Wynter, 2011 |
| | Kalman Filters | – | Lippi et al., 2013; Tampere and Immers, 2007 |
| Traditional Machine Learning | Probabilistic Graph | – | Furtlehner et al., 2021; Zhu et al., 2016 |
| | Support Vector Machine | – | Rapant et al., 2016; Qi and Ishak, 2014 |
| | Gaussian Process Method | – | Yao et al., 2017; Asif et al., 2014; Wang and Shi, 2013 |
| | ANN | – | Lin et al., 2018 |
| | CNN | – | Csikós et al., 2015; Tang et al., 2017 |
| Deep Learning | GRU | – | Zhao et al., 2019b; Chen et al., 2021 |
| | LSTM | – | Cui et al., 2018; Ma et al., 2015 |
| | RNN | SAE (Weather Condition, Time Information & Traffic Accident) | Zhao et al., 2019a; Wang et al., 2019 |
| | CNN | – | Yu et al., 2017; Qu et al., 2021 |
| | Gated CNN | – | |
| | LSTM | – | |
| | LSTM | Weather Condition & Time Information | |
| | LSTM | MLP (Weather Condition & Road Feature) | Yang et al., 2020 |
| | GRU + attention | – | Wu et al., 2018 |
| CapsNet | – | – | Kim et al., 2018 |
| SGCN | CNN | – | Ma et al., 2021 |
| | LSTM | – | |
| | LSTM | Traffic Incidents & Time Information | Xie et al., 2019a |
| | TCN | Road Feature & POI | Ge et al., 2019 |
| | GRU | – | Diao et al., 2019; Zhao et al., 2020b |
| Seq2Seq (RNN)+Attention | – | – | Zhang et al., 2019 |
| DGCN | RNN | – | Chen et al., 2019 |
| | Seq2Seq (GRU) | – | Li et al., 2018b |
| CNN + Attention | The Other Traffic Properties | Feng et al., 2020 |
| | TCN | – | Wu et al., 2019 |
| Attention | Attention | – | Park et al., 2020; Yin et al., 2021b; Zheng et al., 2020 |
| | RNN | – | Zhang et al., 2018 |
| GCN + attention | RNN | – | Huang et al., 2020; Chen et al., 2020a |
| | CNN | – | Zhang et al., 2021a |
| CNN + Attention | RNN | – | Wu et al., 2018 |
| | CNN + Errorfeedback | – | Wang et al., 2016 |
| | RGCN | – | Wang et al., 2018; Guo et al., 2021 |
| GAT + TCN + Attention | – | – | Zhang et al., 2020a |
| Road Network Topology + RNN | – | – | Wang et al., 2019; Kim et al., 2019; Lee et al., 2020 |
The microscopic methods are computationally intensive due to detailed simulation, but they can describe the traffic in detail. On the contrary, macroscopic methods can effectively simulate road networks, but they are challenging to indicate participants’ responses. However, the mesoscopic method combines different advantages of both macroscopic and microscopic methods (Chiu et al., 2010). A representative mesoscopic method is the gas-kinetic model (Helbing, 1996), which captures the dynamic vehicle behavior by probability distributions of traffic properties. Thanks to its mesoscopic level, such methods can achieve a better balance between efficiency and accuracy than the above methods.

In conclusion, model-based methods are difficult to guarantee the real-time requirement of traffic-speed prediction due to the complex traffic model, especially for the microscopic methods, even though they can provide interpretable results. Moreover, artificial assumptions and limited expertise can lead to inevitable errors in the face of dynamic traffic.

Classical data-driven methods

Data-driven methods aim to inductively identify the traffic pattern behind the data and achieve prediction depending on the pattern. The statistical methods predict the future traffic speed primarily based on the temporal dependency of historical speed, of which Autoregressive Integrated Moving Average (ARIMA) is the most popular. ARIMA models the traffic speed as time series and combines autoregressive models, moving average models, and differentiation to capture the temporal dependency. Subsequently, more influencing factors are considered into ARIMA to improve accuracy. For example, SARIMA (Williams and Hoel, 2003) integrated seasonal feature into prediction, and VARIMA (Chandra and Al-Deek, 2009) considered spatial dependency features through upstream and downstream multi-points data. Further, STARIMA (Min and Wynter, 2011) took the limited spatio-temporal feature into account. However, ARIMA and its variants are computationally intensive as the number of relevant points increases. On the other hand, the Kalman filters (KF) are applied to traffic-speed prediction because they can estimate the dynamic system state from noisy data with good real-time performance (Lippi et al., 2013; Tampere and Immers, 2007). Moreover, other statistical methods, such as non-parametric regression (Clark, 2003) and partial least square (Li et al., 2020) are also employed in speed prediction. In addition, the different traffic properties are the contextual factors often considered by statistical methods. Reference (Yu et al., 2019b) used piece-wise function to capture the influence of traffic properties, such as traffic flow of adjacent roads. Meanwhile, the Jenks clustering with dynamic programming are adopted to determine the segment intervals. However, the statistical methods are too simple to extract the dynamic traffic patterns due to the inability to process various factors and stationary assumptions.

Therefore, traditional machine learning is employed in traffic-speed prediction, which can process high-dimensional information and extract complex traffic patterns, and it can be broadly classified into four categories.

Probabilistic Graph Method

Probabilistic Graph Method is the first category, which utilizes a graph to represent the joint probability distribution of the variables in the model. Markov Field is based on undirected acyclic graph (Furtlechner et al., 2021) and static Bayesian Networks (BN) is based on directed acyclic graph (Zhu et al., 2016). Besides, dynamic Bayesian networks are the basis of Hidden Markov Model (HMM) (Rapant et al., 2016). The probabilistic graph can effectively capture the traffic uncertainty, which is an obvious advantage of such methods. Reference (Zhu et al., 2016) integrated the speed categorical variable with spatio-temporal dependency modeling to improve the accuracy through BN because this method can consider continuous and discrete variables together. In (Rapant et al., 2016) and (Qi and Ishak, 2014), the HMM was applied to the speed prediction in freeways, where the dynamic changes of speed states are described by the transition probabilities, and its great robustness for noisy data was shown.

Support Vector Machine

Support Vector Machine (SVM) is the second category, which transforms the input space to a feature space by basis functions and applies a linear model in the feature space (Yao et al., 2017). SVM are binary linear classifiers by drawing the linear boundary in the feature space, and it aims to maximize the gap width. Therefore, SVM is essentially a problem of convex quadratic programming such that global optimization can be reached. Moreover, the nearest data point to the classification boundary is the support vector,
which determines the computational complexity of SVM. In addition, the kernel function ensures the adaptability of SVM for non-linear problems, which transforms the inseparable data in the input space into separable data in the feature space, but the choice of kernel function is still a challenge.

SVM is widely used to extract spatio-temporal features, even the complex features of large-scale networks with different areas (e.g., urban and rural) (Asif et al., 2014). With regard to the kernel function, the common choice is standard functions, such as the Polynomial and Gauss radial basis kernel function, but it is necessary to redesign and adapt to the specific problem. For example, the wavelet function is employed to capture the non-stationary features of traffic speed in a novel kernel function (Wang and Shi, 2013).

Gaussian Process Method

Gaussian Process Method is the third category, which is a generalization of multivariate normal distributions in infinite-dimensional space. The relationship between \( f(X) \) and \( X \) can be expressed as a Gaussian process prior with the mean \( m(X) \) and the covariance function \( k(x, x') \). Besides, the kernel function is the core of a Gaussian process which measures the distance between two sample points. The probability interpretation and CIs of output can be provided by non-parametric Bayesian formulation (Le et al., 2017) in Gaussian Process.

With regard to traffic-speed prediction, this method is mostly employed in the prediction with traffic information fusion, because it can take high-dimensional data, data heterogeneity, uncertainty, and ambiguity into account. For example, the weather information and traffic properties were considered in prediction through Gaussian process (Chen et al., 2014), and the social media data and car trajectory data were taken into account in the same way in (Lin et al., 2018). However, the cubic learning computation and quadratic space requirement are the major limitations of the Gaussian process (Le et al., 2017).

Artificial neural network

Artificial neural network (ANN) is a weighted computational network consisting of several layers of neurons (computational cells), which is inspired by the neural in the biological brain (Csikós et al., 2015) and has been a popular method in traffic-speed prediction thanks to the great robustness. Research (Csikós et al., 2015) used ANN to predict the traffic speed of urban networks. Moreover, a fuzzy ANN was proposed in (Tang et al., 2017), which combined the advantages of fuzzy inference and neural network: knowledge expression and learning ability. Research (Huang and Ran, 2003) inputted the adverse weather information into ANN to consider the impact of weather factor.

Deep-learning methods

Compared to traditional machine learning with shallow structures, deep learning methods can not only handle large-scale data but can also extract the complex pattern between multiple factors and traffic speed. This section discusses the deep learning methods in terms of spatial dependency, temporal dependency, spatio-temporal dependency, and external factors modeling. Based on the generational classification of deep-learning methods for traffic prediction in (Lee et al., 2021), this article does not focus on the first generation methods (i.e., Deep Belief Networks and Stacked AutoEncoder (SAE)) but pays attention to the latest deep learning methods.

Spatial dependency modeling of traffic-speed

CNN. The Convolutional Neural Networks (CNN) are a series of models to process images inspired by the human visual nervous. The convolutional layer applies several convolutional kernels to extract different local features of the image. The pooling layer further reduces the data dimensionality and captures the spatial dependency of different local features. Finally, the fully connected layers output the results based on the features. Moreover, parallel computation is a crucial advantage of CNN thanks to the independence of convolution operations, which can facilitate the training significantly with the weight-sharing mechanism. The traffic speed data in road networks can be modeled as an image and learned by CNN (Tedjopurnomo et al., 2020), and the spatio-temporal matrix was usually used to convert traffic to image (Ma et al., 2017; Wang et al., 2016), where the matrix elements represent speed values in the corresponding position and time.
Thus, it is suitable for spatial dependency modeling and outperforms traditional machine learning even Recurrent Neural Networks (RNN) (Ma et al., 2017). Moreover, researchers have proposed some variants to consider more influencing factors than CNN to improve prediction performance. Reference (Wang et al., 2016) introduced the individual neurons with error feedback mechanisms into CNN to cope with the challenges arising from traffic emergencies, such as traffic peaks and accidents. Concerning long-term prediction, research (Zang et al., 2019) fed the traffic data with different time scales to prediction model and used four CNN modules to extract multiscale spatial features.

**CapsNet.** Despite these considerable advantages of CNN, the pooling operation in CNN results in some important information (i.e., feature locations and their relative spatial relationships) loss and ridiculous results. For example, CNN usually recognizes a human faces with misplaced face features as a normal face. Therefore, the Capsule Network (CapsNet) has recently received a large amount of attention, which replaces the pooling operation with dynamic routing and extracts more accurate spatial relationships among the road segments (Kim et al., 2018). Considering the interrelationship of topology, CapsNet is suitable for large-scale prediction. Reference (Kim et al., 2018) adopted CapsNet to capture the interrelationship of traffic networks in different time steps, which showed a better result than CNN. Reference (Ma et al., 2021) integrated various spatial features (e.g., position, direction, length) of traffic networks into the capsule vectors and extracted comprehensive spatial dependency. However, the training of CapsNet consumes more time than CNN due to its more complex construct.

**GCN.** The traffic speed is described as an image structure to extract spatial dependency in CNNs, but the spatial structure of the traffic network is non-Euclidean in essence, where the relations between road nodes are different. For example, the number of nodes connected to each junction node is not the same, and the road properties of each node are various. Due to such characteristics, it is not feasible to utilize the same size local convolution kernel for all nodes directly such that the convolution kernel of CNN is no longer suitable for non-Euclidean data. Therefore, Graph Convolutional Network (GCN) designed for non-Euclidean data have attracted extensive interest, which can be roughly classified into two categories.

Spectral GCN (SGCN), as the first category, defines the convolution in the spectral domain and transfer traffic data to this domain by graph Fourier transforms (Bruna et al., 2014). Because SGCN can simplify the computation and transform the convolution in spatial domain to a product in frequency domain. From the perspective of signal processing, the graph convolution is regarded as a noise filter (Yin et al., 2021a). Research (Zhao et al., 2020b) used the filter to capture spatial features between the nodes by its first-order neighborhood, then built GCN by stacking multiple convolutional layers. SGCN investigated the graph structures through the eigenvalues and eigenvectors of the Laplacian matrix, which is the difference between the diagonal matrix and the adjacency matrix (Diao et al., 2019). The adjacency matrix of graph structure determines the performance of GCN such that it is one of the research focus. Most existing studies use a single measure to construct the adjacency matrix, leading to an inadequate description of spatial dependency. Therefore, the Multi-Weight GCN considered specific structural features (i.e., speed limit, distance, and angle) in the adjacency matrix and weighted them to describe dynamic spatial dependency. In addition, research (Diao et al., 2019) designed a deep-learning Laplacian matrix estimator to update the real-time matrix and capture the change of traffic pattern.

The other one defines graph convolution based on the spatial structure between graph nodes (Atwood and Towsley, 2016). The diffusion GCN (DGCN) is the most popular that regards graph convolution operation as a diffusion process. Such methods assume the information is transmitted from one node to adjacent nodes with a transition probability such that it can reach a dynamical equilibrium after a few steps. Based on the assumption of state transition, DGCN models the spatial dependency as a stochastic dynamic process, unlike the fixed structure of SGCN. In addition, it generally adopts the bidirectional diffusion mechanism to capture the complex spatial dependency. Research (Li et al., 2018b) integrated DGCN into the Sequence to Sequence (Seq2Seq) to extract the spatial dependency in multi-step speed prediction. To improve the model adaptability for different scenario, research (Wu et al., 2019) proposed a self-adaptive adjacency matrix that can learn unknown graph structures automatically, and the multiple spatial layers were stacked to capture the spatial dependency at different temporal levels.
**Attention.** The attention mechanism is inspired by the attention of human vision, and its core is a set of attention allocation coefficients. In respect of spatial dependency, the attention mechanism can explicitly highlight the roads with a high impact on speed by learning different road unit weights. The most popular work is graph attention networks (GAT) (Pan et al., 2019; Chen et al., 2020a; Huang et al., 2020), which aggregates features from neighboring nodes to the central node by the attention coefficients. Study (Zheng et al., 2020) adopted an attention mechanism to extract temporal and spatial dependency, and proposed a gated method to fuse the features from the attention model (Zheng et al., 2020). Moreover, multi-head attention can capture the features in different subspaces and improve the prediction accuracy, which has lately received great attention. However, traditional multi-head attention treats each head equally. Research (Zhang et al., 2018) further designed a convolutional sub-network to learn the weights and improve the effect of attention. Considering the dynamical change of spatial dependency, the spatial attention module with sentinel vectors are introduced in (Park et al., 2020), which can dynamically adjust the spatial features based on road states. Research (Yin et al., 2021b) proposed an attention mechanism to model the coupling correlations among heterogeneous data. Meanwhile, it divided traffic nodes into different sets based on their adjacency relations and capture the dynamics of different nodes, even global spatial dependency. In respect of data security, research (Zhang et al., 2021a) proposed a differential privacy-based approach and its aggregation mechanism for adjacency matrix, which achieves the trade-off between privacy and performance.

**Temporal dependency modeling of traffic-speed RNN.** RNNs are a class of learning networks designed for sequential data. Different from ANN, each calculation step of RNNs will return the result as an input, and the result of the next step is determined by the new input and previous result. Therefore, RNNs are able to consider the influence of the previous input on the following output with the help of this circular structure. The common format of traffic data is sequential time series, thus RNNs are the ideal candidate for extracting temporal dependency. However, the classical RNN suffers from the attenuation or explosion of network gradients caused by the gradual decrease or increase of parameters in the cyclic computation, leading to the inability for long-term memory in sequences. Moreover, the optimal time lag of RNN needs to be predetermined and mainly is based on a trial-and-error approach.

In an effort to overcome this challenge, researchers proposed Long Short Term Memory Networks (LSTM) by adding a gate control unit to RNN hidden layer. The gates of LSTM are forget, input, and output gates, which selects and store the important information in each step to consider long-term information. As a result, LSTM has been applied more often in speed prediction compared to RNN (Shin and Yoon, 2020; Ma et al., 2015; Cui et al., 2020; Xie et al., 2019a; Niu et al., 2019; Bogaerts et al., 2020). Research (Zang et al., 2019) focused on the long-term prediction at least 24 h and stacked three ConvLSTMs to extract the multiscale temporal features. In addition, the bidirectional mechanism is a significant improvement of classical LSTM (Cui et al., 2018; Wang et al., 2019), which considers the temporal features in both forward and backward directions. Moreover, data quality directly impacts prediction accuracy. A data cleaning rule (Zhao et al., 2019a) and a filling method of missing data based on trend-historical data was adopted before the prediction by LSTM. Research (Tang et al., 2020) constructed estimations for missing values before the predicting by LSTM, which capture global temporal dynamics for missing data and adversarial training to enhance the modeling of global temporal distribution.

Although LSTM can overcome the problem of classical RNN, its three gates structure leads to high computational cost. Therefore, a simple RNN variant, Gate Recurrent Unit (GRU), receives great attention, which simplifies the gate control units of LSTM to two gates (reset gate, update gate) and can achieve similar performance to LSTM in practice (Zhao et al., 2020b; Li et al., 2018b). The bidirectional mechanism can be also integrated with GRU, and research (Chen et al., 2021) adopted the bidirectional GRU to extract the multi-scale temporal dependency based on the multiscale-grid model. Moreover, parameter optimization can directly affect prediction performance. Reference (Zhao et al., 2019b) proposed a GRU with weight optimization for urban express, and showed the great performance of Rmsprop algorithm. Meanwhile, Bayesian optimization is regarded as an effective approach for parameter optimization in (Zhang et al., 2020b). In addition, the Encoder-Decoder structure, usually consisting of the RNNs, is employed to handle the issue of different lengths of input and output sequences (Li et al., 2018b; Zhang et al., 2019, 2020d). However, the limitation of Encoder-Decoder is obvious. No matter what the length of the input and output, the
intermediate vector length is fixed. The issue usually results in a loss of information, and it could become worse as the length of the sequence increases.

**Attention.** The attention mechanism can extract the key parts of inputs rather than encode all the information into vector input. In order to capture dynamic temporal dependency, the attention mechanism has attracted extensive interest (Park et al., 2020). For example, reference (Wu et al., 2018) used the attention model to extract the salient part of traffic data and achieve multi-step prediction. To model the multiscale temporal dependency, research (Feng et al., 2020) designed an attention model to extract whole-range global temporal features and used stacked dilated convolution to capture local temporal features.

**CNN.** Despite RNN is popular in temporal dependency modeling, but it suffers from a high computational cost. On the contrary, CNN can facilitate the training by parallel computation. Thus, research (Diao et al., 2019) captured the short-term temporal features by a 2D temporal convolutional layer. Moreover, the gating mechanism is a common improvement for CNN. The gated linear units’ convolution were employed in both short and long-term prediction (Huang et al., 2020), which simplified the network layers to reduce the accumulated errors in long-term prediction. Research (Liu et al., 2018a) adopted an attention model to weight the feature maps and channels, then extracted temporal features by Gated CNN.

**TCN.** The classical CNN is unable to capture long-term temporal dependency well due to the convolutional kernel size. Therefore, it is generally regarded as an unsuitable method for sequential data compared to RNN. However, CNN has the advantage of computation such that a variant of CNN, Temporal Convolutional Network (TCN), is proposed for modeling sequential data. The size of local input space could affect the convolution result, and is referred to to the receptive field. To expanding the size of receptive field, dilated causal convolution, as the core of TCN, uses the strategy of interval sampling for input, where the receptive field size could grow exponentially with the increase of layer number. Therefore, TCN can obtain a long receptive field with fewer layers, and becomes a candidate for temporal dependency modeling. Besides, TCN is developed from CNN and overcomes the major issue of gradient attenuation or explosion in RNN. The experiment results in (Zhang et al., 2020a) have demonstrated that TCN outperforms RNNs in both accuracy and computation time. Research (Wu et al., 2019) not only integrated gating mechanisms into TCN but also stacked multiple learning layers with TCN to capture the temporal dependency at different temporal levels. Furthermore, the residual connection in TCN allows the model to pass information across layers and reduce the training complexity. Several residual blocks were stacked in TCN to extract the feature of temporal dependency in (Ge et al., 2019) and a suitable number of residual blocks can improve the prediction accuracy.

**Spatio-temporal dependency modeling of traffic-speed**

Most of the existing research extracted the temporal and spatial features separately, and then fuse them by specific methods, such as concatenation, linear transformation, attention mechanism (Xie et al., 2020). The mainstream of spatial dependency modeling is convolutional methods and the popularity of temporal dependency popularity is sequential models. The simple method for capturing spatio-temporal dependency is to concatenate the methods used to extract different features, of which the group of RNN and GCN is the most popular (Wang et al., 2018; Guo et al., 2021). Besides, some researchers proposed an additional module for this problem. Research (Zhang et al., 2020a) used GAT and TCN to extract the two features separately, and further adopted multi-head self-attention to capture the spatio-temporal coupling effects. However, the temporal and spatial features are not independent in essence. The spatial dependency is different at different time steps, and the spatial structure is also an essential factor of temporal dependency. The neglect of interrelationship between the spatial and temporal features is an obvious limitation of the above method, and the deep-learning methods become the promising direction for this issue thanks to its excellent performance in feature extraction. Research (Xie et al., 2020) improved the DGCN to capture the spatio-temporal dependency, where the information was diffused simultaneously to the neighborhood and the next temporal state. In this way, heterogeneous spatial-temporal structures can be modeled as a homogeneous process of diffusion. Another research integrated the topology information of traffic networks into the deep-learning network (Wang et al., 2019) to capture spatio-temporal dependency. In (Kim et al., 2019), the topology of the traffic network was modeled as a spatio-temporal graph and integrated into
RNN to model the complex dependency. Research (Lee et al., 2020) also projected the topology of traffic networks to construct the learning network, and enhanced the model interpretability.

External factors modeling of traffic speed

The spatio-temporal dependency modeling is the focus of the existing works, but it is necessary to consider the external influence factors causing random traffic fluctuations. These factors can be combined with the spatio-temporal learning network or be captured by an additional module. The weather condition and time information are usually considered together. Research (Lv et al., 2018) adopted two fully connected layers to capture the weather and time features. The first layer aimed to extract the holiday and weather features, and the other was used to map features to high dimensions. A Multilayer Perceptron (MLP) was used to capture the features of road properties, weather, and air quality in (Yang et al., 2020), and the traditional traffic theory was applied in the feature merge layer. Reference (Qu et al., 2021) fed the weather and time information as a feature matrix, and an Autoencoder is employed to extract the features in parallel.

The road feature and POI are important factors of spatial dependency. Research (Ge et al., 2019) used POI and the pairwise similarities of graph node to construct the adjacency matrix input. A traffic dataset with various information was released in (Liao et al., 2018), and Seq2Seq was adopted to extract the feature of offline geographical properties, social attributes, and road intersection information.

Urban traffic incidents and social events can reflect the dynamics of the traffic system (Xie et al., 2019a), and such information can be adopted to improve the robustness and adaptability of models. Research (Yu et al., 2017) proposed a DeepLSTM with stacked autoencoder to extract the accident features and jointly model the peak-hour and post-accident conditions traffic. Besides, the impacts of various incidents are different in essence such that research (Xie et al., 2019a) designed a method to discover the traffic incidents with high impact on speed to improve prediction accuracy. Then, a learning network with two fully connected layers was adopted to extract the incident features. In addition, some works integrated traffic properties influence to improve prediction accuracy. Reference (Liu et al., 2018a) constructed the input as 3D matrices by traffic flow, speed, and occupancy to capture the features of traffic properties and improve the robustness for missing data.

Limitation

The deep-learning methods are able to extract diverse traffic features with its superior performance, and have led the research of speed prediction. In the face of the research wave, it is important to make sober and fair assessments of its benefits and drawbacks, and rationally explore the value of such methods.

First, deep-learning methods need to feed large amounts of data, and the data of some important scenarios are difficult to collect, such as accident scenario. Meanwhile, the training cost becomes high due to the heavy data demand and complex model structure. Meanwhile, the computational complexity could increase rapidly as more factors are considered. In addition, due to the black box characteristic, the interpretability of such methods remains a huge issue compared to classical methods. In conclusion, no one method is the panacea, and the deep-learning methods are the right choice only if their unique capabilities are worth its cost (López Manibardo et al., 2021).

PREDICTION METHODS OF VEHICLE SPEED

Prediction methods of vehicle speed can be also divided into model-based, classical data-driven, and deep-learning methods as illustrated in Figure 4. It has to be noticed that vehicle speed is the most uncertain variable at different levels. On the one hand, the vehicle speed is directly affected by its internal factors, and it is difficult to model the complex driving behavior. On the other hand, external factors also impact vehicle speed indirectly. In addition, the spatial dependency of vehicle speed is not very obvious without the networks as a reference. Because of its shortest prediction horizon and real-time requirement among different levels, it is crucial to adopt a less complex model in vehicle-speed prediction. Meanwhile, the adaptability of those prediction methods is more essential than the speed prediction at other levels, because of the various vehicle condition and uncertainty of driving behavior.
The model-based method is interpretable and simple to realize such that it is suitable for vehicle-speed prediction. According to the proposed assumption, model-based methods can be divided into naive and advanced models (Yeon et al., 2019).

Naive model-based methods

The naive model usually used simple rules of vehicle speed or acceleration to describe the vehicle's motion. Such models are only based on real-time vehicle information and can be divided into three classifications: constant speed model, constant acceleration model, and exponential model.

**Constant speed (CS) and constant acceleration (CA) models.**

CS model assumes that the vehicle maintains the same speed while driving,

$$a(t + \Delta t) = 0 \quad \text{(Equation 1)}$$

The CA assumes that the vehicle will maintain the same acceleration,

$$a(t + \Delta t) = a(t) \quad \text{(Equation 2)}$$

The CS model is very popular in collision prediction (Polychronopoulos et al., 2007). It is the basis for calculating the time to collision for risk assessment (Vogel, 2003). CA model, like the CS, is also used in short-term predictions such as collision prediction (Barrios and Motai, 2011). It is also suitable for accurately predicting the motion state of smart devices to achieve positioning functions. Meanwhile, CA and CS models can reliably predict the movement of surrounding vehicles to plan the trajectory of autonomous vehicles (Deo et al., 2018; Wang et al., 2003).

**Exponential models.** Unlike the CA and CS, the exponential model assumes the future speed varies according to a more complex law (i.e. exponential) than CA and CS. Such methods are also based on the current speed information only. Set $t_0$, as the prediction time horizon and $t_1$ as the time step increasing from one to $t_0$. The relationship of the speed of the transition from time $t$ to time $t + t_1$ can be expressed by an exponential function, that is,
\[ v_{t+k} = v_t(1 + \varepsilon)^n, t_k \in 1, 2, \ldots, t_p, \]

(Equation 3)

where \( v_t \) is the speed at time \( t \), \( v_{t+k} \) is the speed obtained by predicting the \( t_k \) second backward at time \( t \), and \( \varepsilon \) is the exponential coefficient.

Furthermore, some researchers assume the torque demand varies according to an exponential law, and the change in torque over time results in fluctuations of speed. Research (Borhan and Vahidi, 2010; Xie et al., 2019b; Zhang et al., 2017b) used the relationship between torque and speed to achieve prediction.

**Advanced model-based methods**

The key of model-based methods is to model the driving behavior, which reflects how the driver response to the environment. Naive models usually only take less internal factors of the vehicle into account and oversimplify the driving behavior. With the development of autonomous vehicles and V2X, advanced models (i.e., car-following model and eco-driving model) have received great attention, which can model a variety of external factors, such as traffic light information, surrounding vehicle information. The car-following model focuses on the interaction with the front vehicle. Based on the speed and position information of the leading vehicle, the intelligent driver model was used in (Jing et al., 2015) to describe the vehicle behavior and estimated the time headway parameter for speed prediction. However, this model only considers the behavior of car-following, which limits the result accuracy, especially in long-term prediction. Therefore, to save vehicle energy in long term, research (Zhou et al., 2019a) planned the vehicles’ trajectories according to the eco-driving rules with the information of traffic lights and surrounding vehicles, and obtained the future speed of ego-vehicle from its trajectory directly. Furthermore, the improvement of accuracy means the increase of computation cost, and the V2X information requires the corresponding devices. Therefore, real-time and information availability should be considered in the advanced model.

**Classical data-driven methods**

The data-driven methods do not require any calibration of the theoretical model and show better accuracy than model-based methods. Classical data-driven methods also can be classified into statistical methods and traditional machine learning methods.

**Classical statistical methods**

Classical statistical methods often are applied to a short-term prediction, mainly including ARIMA and KF. In (Guo et al., 2019b), ARIMA is used to predict vehicle speed for energy management, which is more accurate than many other linear methods. However, the temporal dependency of vehicle-speed prediction is too complex to be effectively captured by the ARIMA model (Liu et al., 2019). Meanwhile, these methods show poor adaptability for a dynamic driving environment. Therefore, hybrid statistical methods receive attention, which uses the complementary of different methods. Research (Qian et al., 2020) used the ARIMA model to predict the vehicle acceleration and input the time-varying result to KF to reach more accurate result. Meanwhile, it proposed an adaptive forgetting factor in KF to improve the filtering accuracy and real-time performance.

**Traditional machine learning**

**Markov.** Such a method can predict the state changes at future moments according to the current state based on Markov theory (Logofet and Lesnaya, 2000). A Markov process is a random process that undergoes a transition from one state to another, and the probability distribution of the next state only depends on the current state instead of the previous sequence (Wang and Infield, 2018). The dynamic driving process can easily be modeled as a Markov process based on the acceleration and speed state transition, and the driving uncertainty can be considered by its simple model structure. Research (Chib, 2001; Sun et al., 2015a) described the driving process by the random Markov model, and uses various cycles to train the state transition matrix. However, the difference between cycles is usually ignored, which could lead to inaccurate state transition and the deviations of prediction result. Therefore, research (Liu et al., 2017) used the neural network to choose the state transition matrix in different driving cycles before the random Markov method. Fuzzy encoding was also used to divide the acceleration events into multiple states (Jing et al., 2017), and the acceleration states are predicted by Markov transition. Besides, the increase of available states in Markov method brings the curse of dimensionality such that some research (Xie et al., 2017) used Monte Carlo approach to obtain the possible future state of Markov chain by sampling in the stationary distribution. Furthermore, it is difficult for a single Markov chain model to use multiple inputs, which
limits its accuracy, such that multiple and self-learning Markov chain models receive the attention. Research (Liu et al., 2021) used historical traffic data to predict a random Markov transition probability, and adopted a neural network to learn the current speed information for prediction. Compared to the traditional Markov model, the self-learning Markov model does not rely on the offline training database to estimate the transition probability matrix, which improves its adaptability under different driving scenarios (Zhou et al., 2019b).

**ANN.** ANN is a popular method for vehicle-speed prediction, especially for long-term prediction (Lefevre et al., 2014), which usually outperforms the Markov and model-based method (Sun et al., 2015a). Meanwhile, the shallow structures of ANN become its advantage compared to the deep-learning methods in vehicle-speed prediction such that a trade-off between accuracy and computation cost can be reached. The first type of ANN is Radial Basis Function neural network (RBF-NN) (Sun et al., 2015b), which can approximate optimal functions with a faster convergence rate than the classical Back Propagation neural network (BP-NN). Research (Sun et al., 2017) inputted the information of driving conditions into RBF-NN and improves the model adaptability under different scenarios. Besides, to capture the speed dynamics, the non-linear autoregressive (NAR) neural network also is employed in vehicle-speed thanks to its dynamic characteristic, which can automatically integrate the output into the input to calculate the next output. Considering the impact of driving behavior on vehicle speed, research (Xiang et al., 2017) used the stroke of accelerator and brake pedal to identify current driving behavior first, and research (Lian et al., 2017) adopted fuzzy inference to identify driving behavior. Then they employed NAR-NN to predict speed under different driving behavior. Furthermore, other methods are usually integrated with ANN to reach an accurate result. Research (Zhang et al., 2020c) proposed a hybrid method consisting of the Markov method and BP-NN, which adopted the Markov method to grasp speed trend and helped BP-NN overcome the local optimal solution.

**Integration with traffic-speed prediction.** Although the definition of speed at different levels is not the same, all of them aim to describe the traffic, just with different perspectives. Speed prediction at different levels present close connections such that it is meaningful to combine them to improve prediction accuracy. The prediction results of traffic-speed are usually regarded as an important reference for vehicle-speed prediction, which reflect the dynamic traffic at micro level (Park et al., 2011). Research (Suh et al., 2020) first employed traffic flow model to predict the traffic state, and the results were used to extract the future position and speed of target vehicle. Analogously, research (Jiang and Fei, 2017) first utilized ANN to predict the traffic speed of road segments, and then adopted HMM to capture the statistical relationship between speed at macro and micro level.

**Deep-learning methods**

Deep-learning methods can capture the complex patterns behind big data and usually obtain more accurate results than the other methods. Meanwhile, the disadvantages, such as high computing and storage cost, are prominent under the real-time requirement of vehicle-speed prediction. However, deep-learning methods are still the trend of vehicle-speed prediction, and it is necessary to continue further research. Unlike the spatio-temporal dependency of traffic speed, temporal dependency is the main characteristic of vehicle speed. Therefore, LSTM, as the mainstream of temporal dependency modeling, is suitable for vehicle-speed prediction. Reference (Liu et al., 2019; Gaikwad et al., 2020) has shown that LSTM can increase the result accuracy and decrease the transient time lag of this prediction compared to the above methods. Research (Yeon et al., 2019) used LSTM to capture the impact of ahead vehicle information, ego-vehicle states and location. This research also showed that the internal vehicle states, radar sensor information, and ego-vehicle location are the essential factors for vehicle speed prediction. Research (Li et al., 2019b) used the Pearson correlation coefficient method to discover the factors with high impact, and then the LSTM was combined with BP to model the temporal dependency. Besides, to especially capture the impact of driving behavior, research (Shih et al., 2019) proposed a hybrid model consisting of coders, LSTM, and attention model. Among them, the attention model is used to identify the start of the driving behavior. Then the especial seq2seq prediction model for different driving behaviors is selected to obtain the future vehicle speed. In addition, CNN is also employed in vehicle-speed prediction, and research (Zhang et al., 2017a) adopted CNN to predict the vehicle speed in different time horizons with the information of the leading vehicle (i.e., its speed and distance to the next traffic light).
LANE-LEVEL SPEED PREDICTION
Every vehicle running on the road is constricted by lane boundaries. The interaction of traffic flow between lanes significantly affects the speed of each vehicle. It is necessary to provide the lane-level speed information for each lane under the perspective of meso. The research on lane-level speed prediction has received attention only in recent years. Therefore, this section primarily elaborates the description and meaning of lane-level speed prediction and reviews the prediction methods at the meso level.

Description of lane-level speed prediction
Compared to the traffic-speed prediction, the lane-level speed prediction releases the assumption that the prediction methods of traffic-speed take the same traffic pattern for multiple lanes on the road. On the one hand, different lanes of the same road have different functions and restrictions thanks to their different locations. For example, the lanes on the China highway are called the overtaking lanes, carriage lanes, and emergency lanes from inside to outside. The speed limit of overtaking lanes is the highest, the occupancy of emergency lanes usually is prohibited. Those characteristics finally result in different lane traffic patterns. On the other hand, because of their interaction, the lane traffic patterns are correlative with the patterns of the other lane on the same road. The real traffic data can demonstrate the difference in lane-level speed. As shown in Figure 5, this article represents the lane-level speed during a day and a week before the overpass and merging lane. The data source is the dataset PeMS (Chen et al., 2001), and the sampling time of the week-speed data is 12:00 of each day. The observation points of the two scenes, which were marked by the dotted frame in the figure, detected the average speed of different lanes simultaneously. Taking scenario A as an example, there were three lanes at the observation point, i.e., lane 1, 2, and 3 as shown in Figure 3. The difference in lane-level speed between them can be up to 15.6% for a day before the overpass and 28.9% for a week before the merging lane. Meanwhile, the trend of different lane-level speed is similar. Therefore, the traffic patterns of different lanes on a specific road are different but correlative with the neighbor.

Lane-level speed prediction can capture different traffic patterns between different lanes and describe vehicle behaviors at meso level (Ke et al., 2020a). Therefore, the spatio-temporal features of lane-level traffic and interaction between lanes bring challenges and opportunities to this prediction. With regard to the application of lane-level speed prediction, it not only provides fine granular future traffic states for traffic management, but also helps the vehicle select a suitable lane and plan the optimal driving path.

Prediction methods of lane-level speed
Lane-level speed prediction is a new research field as compared to traffic and vehicle-speed prediction. Nowadays, the deep-learning method has become the mainstream of speed prediction. Therefore, the
existing research on lane-level speed prediction focuses on this method. The lane-level traffic speed is more difficult to process than the speed at other levels due to the dimension of lane. To simplify the problem, a popular method is to divide the data according to the time or lane. Research (Raza and Zhong, 2017) considered the peak or off-peak conditions and utilized the genetic algorithm to optimize the prediction parameters. Then, the ANN and local weighted regression model are adopted to predict the lane-level speed in different conditions. Research (Tao et al., 2020) used an attention model to extract the important feature and assigned the weight according to different lanes and times. In addition, the multi-channel spatio-temporal image are usually employed for lane-level speed, which can represent the information of temporal and spatial dimension along the corridor as well as the space information across lanes. Processing image data is the strength of CNN such that CNN (Ke et al., 2018) is the preference for this prediction with multi-channel image input. Besides, the characteristic of traffic properties are similar such that the multi-channel image are able to consider the impact of traffic properties. The model construction (Ke et al., 2020b) is similar to that in (Ke et al., 2018), and it converted the lane-level speed and flow data into a spatial-temporal matrix with multi-channel to input CNN.

With regard to modeling the spatial-temporal dependency at lane-level, the common approach is to extract the lane-level temporal and spatial features separately (Ke et al., 2018, 2020b; Raza and Zhong, 2017; Tao et al., 2020), but it ignored the inherent relationship between temporal and spatial features. Research (Lu et al., 2020a) introduced the Conv-LSTM structure to extract spatio-temporal dependency efficiently and simultaneously, and the multi-channel image is suitable to describe the spatio-temporal dependency. Moreover, it is crucial to consider the trade-off between efficiency and accuracy of prediction. Therefore, some research designs a module to discover the feature with high impact before prediction. Research (Lu et al., 2020c) adopted random forest method to extract the importance of the temporal dependency and simplify the input, and then used clockwork RNN to capture the temporal features. Besides, study (Gu et al., 2019b) extracted the spatial features by entropy-based gray relational analysis and selected the lane section with the greatest impact by the correlation analysis methods. The training process in (Gu et al., 2019b; Lu et al., 2020c) was accelerated in the time dimension and space dimension, respectively.

EVALUATION

This section begins with the evaluation metrics of speed prediction and presents the public datasets and open-source codes. Finally, the challenges and future directions of speed prediction are discussed in detail.

Evaluation metrics

It is necessary to evaluate the performance of the speed prediction model for comparing and developing them. The available metrics can be summarized as four categories.

Prediction horizon

The time horizon of prediction, as an important prerequisite, directly determines the characteristics of prediction problem. Meanwhile, the predictability and accuracy generally decrease with the increasing of prediction horizon (Yue et al., 2007).

Absolute error metrics

To evaluate the prediction accuracy, researchers have proposed a series of metrics based on the absolute error between the prediction result and ground truth. Mean absolute error (MAE) and Mean absolute percentage error (MAPE) aim to evaluate the relative bias, and Variance APE (VAPE) denotes the dispersion of absolute error (Yang et al., 2017).

\[
MAPE = \frac{100\%}{MN} \sum_{j=1}^{M} \sum_{i=1}^{N} \frac{y_{ij} - \hat{y}_{ij}}{y_{ij}},
\]  
\[
VAPE = 100\% \times \text{Var} \left( \frac{y - \hat{y}}{y} \right).
\]  
(Equation 4)
(Equation 5)
where \( M \) is the number of time steps, \( N \) is the number of roads, \( \hat{y}_j^i \), \( y_j^i \), and \( u_j^{i_s} \) respectively represent the prediction result, real value, weight of the \( j \)th time step in the \( i \)th road segment, and the vector of \( y_j^i \) and \( y_j^i \) is \( \overline{Y} \) and \( Y \).

**Square error metrics**
Unlike the absolute error metrics, square error metrics use the error square to avoid the complex calculation of absolute values, i.e., mean square error (MSE), root MSE (RMSE), Normalized RMSE, and \( R^2 \).

\[
RMSE = \sqrt{\frac{1}{MN} \sum_{j=1}^{M} \sum_{i=1}^{N} u_j^{i_s} (\hat{y}_j^i - y_j^i)^2},
\]

(Equation 6)

\[
R^2 = 1 - \frac{\sum_{j=1}^{M} \sum_{i=1}^{N} (\hat{y}_j^i - y_j^i)^2}{\sum_{j=1}^{M} \sum_{i=1}^{N} (\overline{Y} - y_j^i)^2},
\]

(Equation 7)

where \( u_j^{i_s} \) is the weight of the \( j \)th time step in the \( i \)th road segment in RMSE, and \( \overline{Y} \) represents the average of \( Y \). A smaller MAPE or RMSE means better prediction accuracy, and a larger \( R^2 \) represents better prediction effect.

**Computation time**
Based on the training process of deep-learning methods, the computation time can be divided into two parts: inference time and training time. The former refers to the computation time to complete a prediction. The longer the inference time the less valuable it is for online prediction. Besides, the training time reflects the complexity of prediction model, and the light model with accuracy result is the trend of development. Furthermore, since the computation time is related to hardware, it makes sense to compare the computation time under the same hardware conditions.

**Public datasets and open-source codes**
Dataset is essential to development and evaluation of prediction models. Table 3 summarizes the public dataset or standard cycle of vehicles applied in different speed predictions. Open-source models can help researchers quickly understand the ideas of existing models and reproduce their methods for comparison. Table 4 provides a list of the public model in different speed predictions to facilitate further research.

**Challenges and future directions**
This section points out the existing challenges and future research opportunities of speed prediction in transportation system as shown in Figure 6.

**Speed-prediction system**

**Systematic speed prediction.** Systematic speed prediction aims to combine the speed predictions at different levels, unlike the speed prediction at a single level, to predict the speed at different levels. The different speed predictions can extract the speed patterns at different levels and promote the understanding of dynamic traffic from macro to micro. First, systematic speed prediction can enhance the deterministic part of the ground truth and extend applications with the help of the information at different levels, especially for Intelligent Vehicle Infrastructure Cooperative Systems (Zhang et al., 2021b). Traffic-speed prediction has been combined with vehicle-speed prediction to improve the prediction performance in (Shao and Sun, 2021; Jiang and Fei, 2017), but the further application of systematic speed prediction is still insufficient and encouraged. Second, a comprehensive understanding of traffic can promote further model development and drive them forward. Besides, current systematic exploration can serve as the basis for the speed prediction of future traffic participants, such as free crowd flow, two-wheeled vehicles, intelligent vehicles, and unmanned air vehicles moving in three dimensions.

**Large-scale traffic data.** The various spatio-temporal data in traffic makes systematic speed prediction feasible. However, large-scale data brings enormous challenge on data processing. Thus, edge computing is regarded as the future framework of speed prediction. Edge computing has been applied to predict the urban energy in (Lee et al., 2019; Luo et al., 2019), but it is necessary to continue further research on edge computing in systematic speed prediction. In addition, the traffic data involves personal privacy and
| Dataset        | Application | Resolution | Location                | Link                                           | Reference                                                                 |
|---------------|-------------|------------|-------------------------|-----------------------------------------------|---------------------------------------------------------------------------|
| PeMSD4        | Traffic     | 5 min      | California, USA         | http://pems.dot.ca.gov/                       | Huang et al., 2020; Yin et al., 2021b; Ge et al., 2019                     |
| PeMSD7        | Traffic     | 5 min      | California, USA         | http://pems.dot.ca.gov/                       | Yu et al., 2019a; Zhang et al., 2021a; Huang et al., 2020; Ge et al., 2019; Yu et al., 2018 |
| PeMSD8        | Traffic     | 5 min      | California, USA         | http://pems.dot.ca.gov/                       | Huang et al., 2020; Yin et al., 2021b                                     |
| PeMS-BAY      | Traffic     | 5 min      | California, USA         | http://pems.dot.ca.gov/                       | Li et al., 2018b; Feng et al., 2020; Chen et al., 2019; Park et al., 2020; Chen et al., 2020a |
| INRIX         | Traffic     | 5 min      | USA                     | https://inrix.com/                            | Lin et al., 2018; Cui et al., 2018, 2020                                    |
| METR-LA       | Traffic     | 5 min      | Los Angeles, USA        | https://www.metro.net/                        | Li et al., 2018b; Feng et al., 2020; Chen et al., 2019; Park et al., 2020; Pan et al., 2019 |
| DRIVE Net     | Traffic     | 5 min      | Seattle, USA            | http://uwdrive.net/STARLab                    | Ma et al., 2018; Qu et al., 2021                                            |
| Madrid city   | Traffic     | 15 min     | Madrid, Spain           | https://datos.madrid.es/portal/site/egob/     | López Manibardo et al., 2021                                               |
| GAIA          | Traffic     | 2~4s       | Chengdu & Xian, China   | https://outreach.didichuxing.com/research/opendata/en/ | Niu et al., 2019                                                            |
| GCM           | Traffic     | 5 min      | Gary Chicago Milwaukee, USA | http://www.travelmidwest.com/                | Polson and Sokolov,2017                                                     |
| Q-Traffic     | Traffic     | 15 min     | Beijing, China          | https://github.com/JingqingZ/BaiduTraffic#q-traffic-dataset | Liao et al., 2018                                                           |
| PortoTaxi     | Traffic     | 15 s       | Porto, Portugal         | https://www.kaggle.com/c/pkdd-15-predict-taxi-servicetrajectory-i | Tang et al., 2020                                                           |
| Seattle Loop  | Traffic     | 5 min      | Seattle, USA            | https://github.com/zhiyongc/Seattle-Loop-Data | López Manibardo et al., 2021; Cui et al., 2020                               |
| NYC           | Traffic     | 5 min      | New York, USA           | https://www.kaggle.com/c/rlaltap/nyc-real-time-traffic-speed-data-feed | López Manibardo et al., 2021                                                 |
| UDDS          | Vehicle     | 1s         | USA                     | https://www.epa.gov/emission-standards-reference-guide | Liu et al., 2017; Cao et al., 2021; Sun et al., 2017; Liu et al., 2021; Sun et al., 2015a; Yan et al., 2018 |
| OCC           | Vehicle     | 1s         | Los Angeles, USA        | https://www.arb.ca.gov/regact/bus02/appc.doc  | Lian et al., 2017                                                           |
| HWFET         | Vehicle     | 1s         | USA                     | https://www.epa.gov/sites/default/files/2015-10/hwy10hztable.txt | Zhou et al., 2019b; Liu et al., 2021; Sun et al., 2017, 2015a               |
| US06          | Vehicle     | 1s         | USA                     | https://www.epa.gov/sites/default/files/2015-10/us06col.txt | Sun et al., 2017; Liu et al., 2021; Sun et al., 2015a; Yan et al., 2018     |
| JN1015        | Vehicle     | 1s         | Japan                   | https://www.epa.gov/sites/default/files/2015-10/jn1015col.txt | Sun et al., 2017                                                            |
| NYCC          | Vehicle     | 1s         | New York, USA           | https://www.epa.gov/sites/default/files/2015-10/nycccol.txt | Liu et al., 2021                                                            |
| SC03          | Vehicle     | 1s         | USA                     | https://www.epa.gov/sites/default/files/2015-10/sc03col.txt | Liu et al., 2021                                                            |
| NEDC          | Vehicle     | 1s         | Europe                  | https://github.com/dabo248/nedc              | Lian et al., 2017; Liu et al., 2021; Sun et al., 2017, 2015a               |
| NGSIM         | Vehicle     | –          | USA                     | https://data.transportation.gov               | Liu et al., 2018b; Zhou et al., 2019a; Shih et al., 2019; Ye et al., 2019 |
national security, and the huge amount of data means huge information-security issues (Zhang et al., 2021a). Moreover, privacy protection implies a reduction of computation efficiency, and it is essential to consider the trade-off between efficiency and privacy security.

Information basis

Various information. The information and knowledge used for prediction determine the predictable part of the ground truth. The evolution of the prediction model corresponds to the increase of the information used in the model. However, most of the existing studies consider the external factors individually and ignore the coupled relationship between various factors. Thus, it is become important to further extract the complex traffic patterns. Moreover, ITS is a potential way of information collection, which is able to achieve real-time collection across space and enlarge the single-vehicle view. Concerning traffic knowledge, existing methods mostly focus on spatio-temporal dependency. Furthermore, extracting spatio-temporal dependency at a larger scale is still a future direction. The prediction research should be extended from short-term to long-term in the temporal dimension, and from a single vehicle, lane, road segment to the road network in the spatial dimension. Moreover, the joint spatio-temporal dependency is urgently needed in future research. Furthermore, in addition to spatio-temporal dependency, more systematic features of influencing factors should be considered.

Fusion of multiple information. The complementarity of multi-sources information can compensate for missing or noisy data. With regard to the fusion method, feature-level fusion can achieve a better performance than data-level fusion (Li et al., 2019a). However, the characteristics (e.g., type, resolution) of multi-sources data are different, and the process methods of different data are the first problem of information fusion. Furthermore, as the information dimensions increase, the difficulty of feature extraction and computational consumption will rise significantly. Therefore, the balance between data requirement and computational efficiency is an important direction.

Prediction models

Understanding and interpreting models. Nowadays, the mainstream of speed prediction is deep-learning, but it has been criticized for its black-box characteristics. The end-to-end characteristics allow us to reach accuracy results without understanding how to obtain them, which greatly reduces the reliability and transferability of such methods. In practice, one might even prefer the more interpretable but less performant solution (López Manibardo et al., 2021). A potential direction is to map the physical features of information and knowledge into the deep-learning network or input. GCN replaces the input from the image to the graph based on non-Euclidean features of the road network and improves the prediction accuracy. Reference (Wang et al., 2019; Kim et al., 2019; Lee et al., 2020) took advantage of the road network topology to build up the deep-learning network. However, prediction models with good interpretability need to be further explored.

Robustness of models. The robustness of prediction models usually refers to the ability to resist perturbations and noise. During the data collection, data is inevitably missing and noisy due to environmental influences and sensor limitations. Most of the studies separated the data processing (i.e. missing and noise handling) from the model itself (Wang and Shi, 2013; Zhao et al., 2019a; Chen et al., 2020b), and some work

| Application | Model | Reference | Year | Framework | Link |
|-------------|-------|-----------|------|-----------|------|
| Traffic     | DCRNN | Li et al., 2018b | 2018 | Tensorflow | https://github.com/liyaguang/dcrnn |
|             | GRNN  | Wang et al., 2018 | 2018 | –         | https://github.com/xxArbiter/grnn |
|             | STGCN | Yu et al., 2018 | 2018 | keras     | https://github.com/Knowledge-Precipitation-Tribe/STGCN-keras |
|             | Graph WaveNet | Wu et al., 2019 | 2019 | Torch     | https://github.com/nnzhan/Graph-WaveNet |
|             | ST-MetaNet | Pan et al., 2019 | 2019 | MXNet     | https://github.com/panzheyi/ST-MetaNet |
|             | T-GCN  | Zhao et al., 2020b | 2020 | Tensorflow | https://github.com/lehaifeng/T-GCN |
|             | GMAN   | Zheng et al., 2020 | 2020 | Tensorflow | https://github.com/zhengchuanpan/GMAN |
| Lane-level  | MDL    | Lu et al., 2020a | 2020 | –         | https://github.com/lwqs93/MDL |
|             | ST-AFN | Shen et al., 2021 | 2021 | –         | https://github.com/MCyutou/ST-AFN |
treated this part as a new learning task (Tian et al., 2018; Zhang et al., 2021d). Besides, it is a future direction for the development of model robustness to combine data imputation with prediction methods (Zhang et al., 2021d; Chen and Sun, 2021; Boquet et al., 2020). However, data imperfections are inevitable in realistic environments, thus developing predictive models with good robustness requires more effort.

**Adaptability of models.** The adaptability describes the working range of the scenarios prediction methods covered. The first is the adaptability to different structures of road networks. For example, the traffic patterns on urban roads are more complex than that of freeways due to the complex intersection and traffic light control (Jin, 2010), leading to their difference of traffic pattern. The second is the adaptability to different traffic scenarios, especially the rare scenarios, such as the accident scenario. The adaptability to different vehicles (e.g. emergency vehicles (Zhao et al., 2020a) and trucks (Zhao et al., 2019b)) is also significant. In addition, the adaptability to the change of traffic patterns is also worth noticing. Our city is developing all the time, and the reshaping of road leads to the old traffic data being useless, which is defined as concept drift (Lana et al., 2018). Therefore, prediction models should adapt to the new traffic data, and some graph-based prediction methods adopted dynamic input matrix to model the dynamic spatio-temporal dependency (Diao et al., 2019; Zheng et al., 2020). It is necessary to continue further research in this direction. Finally, as data-driven methods do not require explicit modeling, the speed prediction models can also be applied to other prediction and promote the development of prediction method.

**Emerging models.** A number of deep-learning methods have been adopted for speed prediction, such as Transformers, Generative Adversarial Networks (GAN), Meta learning. Transformers (Xu et al., 2021) is based on Encoder-Decoder structure and can achieve parallel computation compared to RNN thanks to the self-attention mechanism. GAN (Yu and Gu, 2019; Zhang et al., 2021c) captures the data distribution by two networks competing with each other and exhibit strong robustness. Meta learning (Pan et al., 2019) possesses self-learning capability to leverage experience to guide future tasks. Even though most of them in speed prediction are still in their infancy, those methods inject new life into speed prediction.

**Selection of the optimal model.** How to choose the right method for a certain problem needs further exploration. No one model is suitable for any problem, and a method is a suitable choice only if it reaches the trade-off between the benefits and cost (López Manibardo et al., 2021). This article attempts to present the existing methods in terms of the utilization of different information, but the real problems require additional considerations such as hardware limitations, real-time performance, data needs, accuracy, etc. How
to balance these requirements to select the right model requires further research. Meanwhile, most of the existing work has focused on the design of new models, but not enough attention has been paid to the model parameter optimization. The potential of existing models needs to be further explored. In addition, the trend of existing models is toward more complex. Most works mechanically stack spatio-temporal layers to extract spatio-temporal dependency. However, considering the practical applications and real-time requirements, light prediction models are also an important direction.

**Standard evaluation systems**

Many exciting methods have emerged from the research wave of the data-driven method, but the development of speed prediction is based on the side-by-side comparison of different methods. However, the existing methods have different experimental datasets and evaluation metrics, which makes it difficult to state whether the performance improvement is thanks to parameter tuning or model improvement. Therefore, a benchmark evaluation system is urgently needed, including standard datasets, prediction models, and reasonable evaluation metrics.

**Benchmark datasets and prediction models.** Benchmark datasets can standardize the experiment and compare the performance of individual methods in the same arena. In addition, benchmark-dataset management can directly address the issues mentioned in Section 6.3.2 by itself, such as the lack of external factors, the inaccuracy of existing datasets, etc. The standard prediction model delineates an evaluation reference for the benchmark evaluation, and reduces the workload of comparison experiments.

**Reasonable standard metrics.** A reasonable evaluation system is essential for identifying model problems and guiding model development. However, the current evaluation mostly focuses on accuracy evaluation. In fact, the various performances of practical application should be evaluated. In addition, the evaluation metrics should not be limited to the result evaluation but can further evaluate the performance of internal system characteristics. For example, reference (Wu et al., 2018) proposed a metric to measure the prediction performance of spatial and temporal distribution. Moreover, the model’s potential should be fully explored to provide other valuable information and enhance the result reliability.

**CONCLUSION**

This study conducted a comprehensive review of speed prediction at different levels in the transportation system. Specifically, this article analyzed the information used and the prediction model on the prediction results. First, the speed at different levels i.e. traffic (macro), vehicle (micro), and lane-level (meso) aims to describe the dynamic traffic, but from different perspectives. Therefore, the main focus of prediction methods is from the spatio-temporal dependency and external factors modeling to temporal dependency and internal factors modeling, especially to that of micro driving behavior. In addition, the article summarized the prediction methods at different levels based on how prediction models use the different information to meet the challenge of different prediction problems. Furthermore, the article reviewed the existing evaluation metrics, public datasets, and open-source prediction models. Finally, the existing challenges and future directions of speed prediction were discussed in detail. To the best of our knowledge, this article is the first review for speed prediction at different levels, and the lane-level speed prediction is the first time to be elaborated in a review. Therefore, this article can help researchers quickly locate their research in this field and find potential branches to further explore. We hope that the discussion of speed prediction with a systematic perspective can probe the pulse of traffic and improve city transportation.

**Limitations of the study**

This article focuses on the systematic review of speed prediction methods at different levels in transportation system. A quantitative comparison of different prediction method in specific scenarios are beyond the reach of this article and require further research.

**ACKNOWLEDGMENTS**

The authors would like to thank the National Key R&D Program of China under Grant NO2020AAA0108100, Science and Technology Commission of Shanghai Municipality under Grant 21ZR1465900, and the anonymous reviewers for their valuable suggestions.
AUTHOR CONTRIBUTIONS

Conceptualization: Y. Huang and Z. Zhou; Writing of Original Draft: Z. Zhou and Z. Yang; Revision: Z. Zhou, Z. Yang, Y. Huang, and Z. Zhang; Supervision: Y. Huang, Z. Zhang, H. Chen, and Z. Yu; Funding Acquisition: Y. Huang, H. Chen, and Z. Yu.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Ahmed, M.M., and Ghasemzadeh, A. (2018). The impacts of heavy rain on speed and headway Behaviors: an investigation using the SHRP2 naturalistic driving study data. Transportation Res. C. Emerging Tech. 91, 371–384. https://doi.org/10.1016/j.trc.2018.04.012.

Amini, M.R., Feng, Y., Yang, Z., Kolmanovsky, I., and Sun, J. (2020). Long-term vehicle speed prediction via historical traffic data analysis for improved energy efficiency of connected electric vehicles. Transportation Res. Rec. J. Transportation Res. Board 2674. 0361198120941508. https://doi.org/10.1017/S0017868819003497.

Asif, M.T., Dauwels, J., Goh, C.Y., Orazi, A., Fathi, E., Xu, M., Dhanya, M.M., Mitrovic, N., and Jallet, P. (2014). Spatiotemporal patterns in large-scale traffic speed prediction. IEEE Trans. Intell. Transportation Syst. 15, 794–804.

Atwood, J., and Towsley, D. (2016). Diffusion-convolutional neural networks. Advances in neural information processing systems 29.

Barrios, C., and Motai, Y. (2011). Improving estimation of vehicle’s trajectory using the latest global positioning system with kalman filtering. IEEE Trans. Instrumentation Meas. 60, 3747–3755. https://doi.org/10.1109/TIM.2011.2147670.

Bogaerts, T., Masegosa, A., Angarita-Zapata, J., Onieva, E., and Hellinckx, P. (2020). A graph CNN-LSTM neural network for short and long-term traffic forecasting based on trajectory data. Transportation Res. C. Emerging Tech. 112, 62–77. https://doi.org/10.1016/j.trc.2020.01.010.

Boquet, G., Morell, A., Serrano, J., and Vicario, J.L. (2020). A variational autoencoder solution for road traffic forecasting systems: missing data imputation, dimension reduction, model selection and anomaly detection. Transportation Res. Part C. Emerging Tech. 115, 102622.

Borhan, H.A., and Vahidi, A. (2010). Model predictive control of a power-split Hybrid Electric Vehicle with combined battery and ultracapacitor energy storage. In Proceedings of the 2010 American Control Conference, IEEE, Baltimore, MD, I, pp. 5031–5036. https://doi.org/10.1109/ACC.2010.5530728.

Bruna, J., Zaremba, W., Salam, A., and LeCun, Y. (2014). Spectral Networks and Locally Connected Networks on Graphs. Preprint at. arXiv. arXiv:1312.6203.

Cao, J., He, H., and Cui, X. (2021). City buses’ future velocity prediction for multiple driving cycle: a meta supervised learning solution. IET Intell. Transport Syst. 15, 359–370. https://doi.org/10.1049/itr2.12019.

Chandra, S.R., and Al-Deek, H. (2009). Predictions of freeway traffic speeds and volumes using vector autoregressive models. J. Intell. Transportation Syst. 13, 53–72.

Chen, C., Li, K., Teo, S., Zhou, X., Wang, K., Wang, J., and Zeng, Z. (2019). Gated residual recurrent graph neural networks for traffic prediction. Proc. AAAI Conf. Artif. Intelligence 33, 485–492.

Chen, C., Petty, K., Skabardonis, A., Vaniaya, P., and Jia, Z. (2011). Freeway performance measurement system: mining loop detector data. Transportation Res. Rec. 1748, 96–102.

Chen, D., Yan, X., Liu, X., Li, S., Wang, L., and Tian, X. (2021). A multiscale-grid-based stacked bidirectional GRU neural network model for predicting traffic speeds of urban expressways. IEEE Access 9, 1321–1337.

Chen, W., Chen, L., Xie, Y., Cao, W., Gao, Y., and Feng, X. (2020a). Multi-range attentive bicomponent graph convolutional network for traffic forecasting. Proc. AAAI Conf. Artif. Intelligence 34, 3529–3536.

Chen, X., and Sun, L. (2021). Bayesian temporal factorization for multidimensional time series prediction. In IEEE Transactions on Pattern Analysis and Machine Intelligence, p. 1.

Chen, X., Wu, S., Shi, C., Huang, Y., Yang, Y., Ke, R., and Zhao, J. (2020b). Sensing data supported traffic flow prediction via denoising schemes and ANN: a comparison. IEEE Sensors J. 20, 14317–14328.

Chen, X.Y., Pao, H.K., and Lee, Y.J. (2014). Adjustment of the parameters of the urban traffic flow prediction. In Conference on Computer Science and Software Engineering, pp. 29–33.

Chen, W., Chen, L., Xie, Y., Cao, W., Gao, Y., and Feng, X. (2020a). Multi-range attentive bicomponent graph convolutional network for traffic forecasting. Proc. AAAI Conf. Artif. Intelligence 34, 3529–3536.

Chen, X., and Sun, L. (2021). Bayesian temporal factorization for multidimensional time series prediction. In IEEE Transactions on Pattern Analysis and Machine Intelligence, p. 1.

Chen, X., Wu, S., Shi, C., Huang, Y., Yang, Y., Ke, R., and Zhao, J. (2020b). Sensing data supported traffic flow prediction via denoising schemes and ANN: a comparison. IEEE Sensors J. 20, 14317–14328.

Chen, X.Y., Pao, H.K., and Lee, Y.J. (2014). Adjustment of the parameters of the urban traffic flow prediction. In Conference on Computer Science and Software Engineering, pp. 29–33.

Chib, S. (2001). Markov chain Monte Carlo methods: computation and inference. In Handbooks in Economics, $\text{\&}$Handbooks in Economics (Elsevier), pp. 3569–3669. https://doi.org/10.1016/S1573-4412(01)05010-3.

Chiu, Y.C., Zhou, L., and Song, H. (2010). Development and calibration of the Anisotropic Mesoscopic Simulation model for uninterrupted flow facilities. Transportation Res. B: Methodological 44, 152–174.

Clark, S. (2003). Traffic prediction using multivariate nonparametric regression. J. Transportation Eng. 129, 161–168.

Csikós, A., Viharos, Z.J., Kis, K.B., Tettamanti, T., and Varga, I. (2015). Traffic speed prediction method for urban networks—an ANN approach. In 2015International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS) (IEEE), pp. 102–108.

Cui, Z., Henrickson, K., Ke, R., and Wang, Y. (2020). Traffic graph convolutional recurrent neural network: a deep learning framework for network-scale traffic learning and forecasting. IEEE Trans. Intell. Transportation Syst. 21, 4883–4894.

Cui, Z., Ke, R., and Wang, Y. (2018). Deep stacked bidirectional and unidirectional LSTM recurrent neural network for network-wide traffic speed prediction. Transportation Res. Part C: Emerging Tech. 118, 102674–102688.

Deo, N., Rangesh, A., and Trivedi, M.M. (2018). How would you be driving in a world with autonomous cars? A unified framework for maneuver classification and motion prediction. IEEE Trans. Intell. Vehiciles 3, 129–140. https://doi.org/10.1109/TIV.2018.2804159.

Diao, Z., Wang, X., Zhang, D., Liu, Y., Xie, K., and He, S. (2019). Dynamic spatial-temporal graph convolutional neural networks for traffic forecasting. Proc. AAAI Conf. Artif. Intelligence 33, 890–897.

Feng, D., Wu, Z., Zhang, J., and Wu, Z. (2020). Dynamic global-local spatial-temporal network for traffic speed prediction. IEEE Access 8, 209296–209307.

Fridman, L., Brown, D.E., Glazer, M., Angell, W., Dodd, S., Jenik, B., Terwilliger, J., Patsekin, A., Kindelsberger, J., Ding, L., et al. (2019). MIT advanced vehicle Technology study: large-scale naturalistic driving study of driver behavior and interaction with automation. IEEE Access 7, 102021–102038. https://doi.org/10.1109/ACCESS.2019.2926040.

Furthner, C., Lasgouttes, J.M., Attanasi, A., Pezzula, M., and Gentile, G. (2021). Short-term forecasting of urban traffic using spatio-temporal Markov field. IEEE Trans. Intell. Transportation Syst. 1–10.

Gaikwad, T., Rabinowitz, A., Motalebiaregha, F., Bradley, T., Asher, Z., Fong, A., and Meyer, R. (2020). Vehicle velocity prediction using artificial neural network and effect of real World signals on prediction window. In WCX SAE World Congress Experience.

Ge, L., Li, H., Liu, J., and Zhou, A. (2019). Temporal graph convolutional networks for traffic speed prediction considering external factors. In 2019 20th IEEE International Conference on Mobile Data Management (MDM), pp. 234–242.

Gu, W., Zhao, D., and Mason, B. (2019a). A review of intelligent road preview methods for energy management of hybrid vehicles. IFAC-PapersOnLine 52, 654–660.
Gu, Y., Lu, W., Qin, L., Li, M., and Shao, Z. (2019a). Short-term prediction of lane-level traffic speeds: a fusion deep learning model. Transportation Res. Part C: Emerging Tech. 106, 1–16. https://doi.org/10.1016/j.trc.2019.07.003.

Guo, J., He, H., Peng, J., and Zhou, N. (2019a). A novel MPC-based adaptive energy management strategy in plug-in hybrid electric vehicles. Energy 173, 378–392. https://doi.org/10.1016/j.energy.2019.03.083.

Guo, J., He, H., and Sun, C. (2019b). ARIMA-based road gradient and vehicle velocity prediction for hybrid electric vehicle energy management. IEEE Trans. Vehicular Technol. 68, 5309–5320. https://doi.org/10.1109/TVT.2019.2921293.

Guo, K., Hu, Y., Qian, Z., Liu, H., Zhang, K., Sun, Y., Gao, J., and Yin, B. (2021). Optimized graph convolution recurrent neural network for traffic prediction. IEEE Trans. Intell. Transportation Syst. 22, 1138–1149. https://doi.org/10.1109/TITS.2019.2921372.

He, H., Wang, Y., Han, R., Han, M., Bai, Y., and Liu, Q. (2021). An improved MPC-based energy management strategy for hybrid vehicles using V2V and V2I communications. Energy 225, 120073. https://doi.org/10.1016/j.energy.2021.120073.

Helbing, D. (1996). Gas-kinetic derivation of Navier-Stokes-like traffic equations. Phys. Rev. E 53, 2366–2381.

Huang, K., Yang, X., Lu, Y., Mi, C.C., and Kondapudi, P. (2018). Ecological driving system for connected/automated vehicles using a two-stage control hierarchy. IEEE Trans. Transportation Syst. 19, 2373–2384.

Huang, R., Huang, C., Liu, Y., Dai, G., and Kong, W. (2020). LSGCN: long short-term traffic prediction with graph convolutional networks. In Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, International Joint Conferences on Artificial Intelligence Organization, Yokohama, Japan, pp. 2355–2361.

Huang, S.H., and Ran, B. (2009). An application of neural network on traffic speed prediction under adverse weather condition. In Transportation Research Board 82nd Annual Meeting/Transportation Research Board, Transportation Research Board., ed. (The University of Wisconsin-Madison), pp. 1–21.

Huang, Y., Wang, H., Khajepour, A., Hongwen, h., and j. (2017). Model predictive control power management strategies for HEVs: a review. J. Power Sourc. 341, 91–106.

Hyeon, E., Kim, Y., Prakash, N., and Stefanopoulou, A. (2019). Short-term traffic forecasting using vehicle wireless communications. In 2019 American Control Conference (ACC), p. 741. https://doi.org/10.23919/ACC.2019.8814400.

Jiang, B., and Fei, Y. (2017). Vehicle speed prediction by two-level data driven models in vehicular networks. IEEE Trans. Intell. Transportation Syst. 18, 1793–1801.

Jin, W.L. (2010). Continuous kinematic wave models of merging traffic flow. Transportation Res. Part B: Methodological 44, 1084–1103.

Jing, J., Filev, D., Kurt, A., Ozatay, E., Michelini, J., and Ozguner, U. (2017). Vehicle speed prediction using a cooperative method of fuzzy Markov model and auto-regressive model. In 2017 IEEE Intelligent Vehicles Symposium (IV), pp. 881–886. https://doi.org/10.1109/IVS.2017.7959327.

Jing, J., Kurt, A., Ozatay, E., Michelini, J., Filev, D., and Ozguner, U. (2015). Vehicle speed prediction in a convoy using V2V communication. In 2015 IEEE 18th International Conference on Intelligent Transportation Systems, pp. 2861–2868.

Ke, R., Feng, S., Cui, Z., and Wang, Y. (2020a). Advanced framework for microscopic and lane-level macroscopic traffic parameters estimation from UAV video. IET Intell. Transport Syst. 14, 724–734. https://doi.org/10.1049/iet-its.2019.0463.

Ke, R., Li, W., Cui, Z., and Wang, Y. (2018). Multi-lane traffic pattern learning and forecasting using convolutional neural network. In COTA International Symposium on Emerging Trend in Transportation, p. 8.

Ke, R., Li, W., Cui, Z., and Wang, Y. (2020b). Two-stream multi-channel convolutional neural network for multi-lane traffic speed prediction considering traffic volume impact. Transportation Res. Rec. 2674, 459–470.

Kim, Y., Wang, P., and Mihaylova, L. (2019). Structural recurrent neural network for traffic speed prediction. In ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5207–5211.

Kim, Y., Wang, P., Zhu, Y., and Mihaylova, L. (2018). A capsule network for traffic speed prediction in complex road networks. In 2018 Sensor Data Fusion: Trends, Solutions, Applications (SDF), pp. 1–6.

Korčěk, P., Sekanina, L., and Fusič, O. (2011). A scalable cellular automata based microscopic traffic simulation. In in 2011 IEEE Intelligent Vehicles Symposium (IV), pp. 13–18.

Lana, I., Ser, J.D., Velez, M., and Vlahogianni, E.I. (2018). Road traffic forecasting: recent advances and new challenges. IEEE Intell. Transportation Syst. Mag. 10, 93–109. https://doi.org/10.1109/MITS.2018.2806634.

Lee, T.V., Oentaryo, R., Liu, S., and Lau, H.C. (2017). Local Gaussian processes for efficient fine-grained traffic speed prediction. IEEE Trans. Big Data 3, 194–207.

Lee, K., Eo, M., Jung, E., Yoon, Y., and Rhee, W. (2021). Short-term traffic prediction with deep neural networks: a survey. IEEE Access 9, 54739–54756.

Lee, S.H., Lee, T., Kim, S., and Park, S. (2019). Energy consumption prediction system based on deep learning with edge computing. In 2019 IEEE 2nd International Conference on Electronics Technology (ICET), pp. 473–477.

Lee, Y., Jeon, H., and Sohn, K. (2020). Predicting short-term traffic speed using a deep neural network to accommodate citywide spatio-temporal correlations. IEEE Transactions on Intelligent Transportation Systems 22, 1435–1448.

Lefevere, S., Sun, C., Bajcsy, R., and Laugier, C. (2014). Comparison of parametric and non-parametric approaches for vehicle speed prediction. In Proceedings of the American Control Conference, pp. 3494–3499. https://doi.org/10.1109/CDC.2014.6858871.

Li, L., Xu, Z., Zhang, J., Wang, Y., and Ran, B. (2019a). Traffic speed prediction for intelligent transportation system based on a deep feature fusion model. J. Intell. Transportation Syst. 23, 605–616.

Li, W., Wang, J., Fan, R., Zhang, Y., Guo, Q., Siddiquee, C., and Ban, X.J. (2020). Short-term traffic state prediction from latent structures: accuracy vs. efficiency. Transportation Res. Part C: Emerging Tech. 111, 72–90.

Li, Y., Chen, M., Lu, X., and Zhao, W. (2018a). Research on optimized GA-SVM vehicle speed prediction model based on driver-vehicle-road-traffic system. Sci. China Technol. Sci. 61, 782–790. https://doi.org/10.1007/s11431-017-9213-0.

Li, Y., Yu, R., Shahabi, C., and Liu, Y. (2018b). Diffusion convolutional recurrent neural network: data-driven traffic forecasting. In International Conference on Learning Representations, pp. 1–16.

Lian, J., Liu, S., Liu, L., Liu, X., Zhou, Y., Yang, F., and Yuan, L. (2017). A mixed logical dynamical-model predictive control (MLD-MPC) energy management control strategy for plug-in hybrid electric vehicles (PHEVs). Energies 10, 74. https://doi.org/10.3390/en10010074.

Liao, B., Zhang, J., Wu, C., Mdlwraith, D., Chen, T., Yang, S., Guo, Y., and Wu, F. (2018). Deep sequence learning with auxiliary information for traffic prediction. In The 24th ACM SIGKDD International Conference, p. 346. https://doi.org/10.1145/3219819.3219895.

Liebig, T., Piatakowski, N., Bockermann, C., and Morik, K. (2017). Dynamic route planning with real-time traffic predictions. Inf. Syst. 64, 258–265.

Lin, L., Li, J., Chen, F., Ye, J., and Huai, J. (2018). Road traffic speed prediction: a probabilistic model fusing multi-source data. IEEE Trans. Knowledge Data Eng. 30, 1310–1323.

Lippi, M., Bertini, M., and Frasconi, P. (2013). Short-term traffic flow forecasting: an experimental comparison of time-series analysis and supervised learning. IEEE Trans. Intell. Transportation Syst. 14, 671–682.

Liu, C.M., and Juang, J.C. (2021). Estimation of lane-level traffic flow using a deep learning technique. Appl. Sci. 11, 5619. https://doi.org/10.3390/app11255619.

Liu, Q., Wang, B., and Zhu, Y. (2018a). Short-term traffic speed forecasting based on attention convolutional neural network for arterials.
Computer-Aided Civil Infrastructure Eng. 33, 999–1016.
Liu, K., Asher, Z., Gong, X., Huang, M., and Kolmanovsky, I. Vehicle Velocity Prediction and Energy Management Strategy Part 1: Deterministic and Stochastic Vehicle Velocity Prediction Using Machine Learning. https://doi.org/10.4271/2019-01-1051.
Liu, Y., Cheng, D., Wang, Y., Cheng, J., and Gao, S. (2018b). A novel method for predicting vehicle state in internet of vehicles. Mobile Inf. Syst. 2018, e9728329. https://doi.org/10.1155/2018/9728328.
Liu, Y., Li, J., Gao, J., Lei, Z., Zhang, Y., and Chen, Z. (2021). Prediction of vehicle driving conditions with incorporation of stochastic forecasting and machine learning and a case study in energy management of plug-in hybrid electric vehicles. Mech. Syst. Signal Process. 158, 107765.
Liu, Y.d., Chu, L., Xu, N., Jia, Y.f., and Xu, Z. (2017). A method to improve accuracy of velocity prediction using H. Markov model. In Neuro Information Processing, 10638. D. Liu, S. Xie, Y. Li, D. Zhao, and E.S.M. El-Alfy, eds. (Springer International Publishing), pp. 378–386. https://doi.org/10.1007/978-3-319-70139-4_38.
Logofet, D.O., and Lesnaya, E.V. (2000). The mathematics of Markov models: what Markov chains can really predict in forest successions. Ecol. Model. 126, 285–298. https://doi.org/10.1016/S0304-3800(00)00269-6.
López Manibardo, E., Laina, I., and Del Ser, J. (2021). Deep learning for road traffic forecasting: does it make a difference? IEEE Transactions on Intelligent Transportation Systems, 1–25. https://doi.org/10.1109/TITS.2021.3083957.
Lu, W., Rui, Y., and Ran, B. (2020a). Lane-level traffic speed forecasting: a novel mixed deep learning model. IEEE Trans. Intell. Transportation System. 1–12.
Lu, W., Rui, Y., Yi, Z., Ran, B., and Gu, Y. (2020b). A hybrid model for lane-level traffic flow forecasting based on complete ensemble empirical mode decomposition and extreme gradient boosting. IEEE Access 8, 42042–42054. https://doi.org/10.1109/ACCESS.2020.2977219.
Lu, W., Yi, Z., Liu, W., Gu, Y., Rui, Y., and Ran, B. (2020c). Efficient deep learning based method for multi-lane speed forecasting: a case study in Beijing. IET Intel. Transport Syst. 14, 2017–2023.
Luo, H., Cai, H., Yu, H., Sun, Y., Bi, Z., and Jiang, L. (2019). A short-term energy prediction system based on edge computing for smart city. Future Generation Computer Syst. 101, 444–457.
Lv, Z., Xu, J., Zheng, K., Yin, H., Zhao, P., and Zhou, X. (2018). LC-RNN: a deep learning model for traffic speed prediction. IJCAI, 3470–3476.
Ma, D., Sheng, B., Jin, S., Ma, X., and Gao, P. (2018). Short-term traffic flow forecasting by selecting appropriate predictions based on pattern matching. IEEE Access 6, 75629–75638.
Ma, X., Dai, Z., He, Z., Ma, J., Wang, Y., and 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.
Ma, X., Tuo, Z., Wang, Y., Yu, H., and Wang, Y. (2015). Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. Transportation Res. Part C: Emerging Tech. 54, 187–197.
Ma, X., Zhong, H., Li, Y., Ma, J., Cui, Z., and Wang, Y. (2021). Forecasting transportation network speed using deep capsule networks with nested LSTM models. IEEE Trans. Intell. Transportation Syst. 22, 4813–4824. https://doi.org/10.1109/TITS.2020.2984813.
Maerivoet, S., and De Moor, B. (2005). Cellular automata models of road traffic. Phys. Rep. 419, 1–64.
Miglani, A., and Kumar, N. (2019). Deep learning models for traffic flow prediction in autonomous vehicles: a review. Sensors. https://doi.org/10.3390/s19081984.
Min, W., and Wynter, L. (2011). Real-time road traffic prediction: a review of mathematical correlations. Transportation Res. Part C: Emerging Tech. 19, 606–616.
Moser, D., Wasch, H., Schmied, R., Efendic, H., and del Re, L. (2015). Short term prediction of a vehicle’s trajectory using ITS. SAE Int. J. Passenger Cars - Electron. Electr. Syst. 8, 364–370.
Nagel, K., and Schreckenberg, M. (1992). A cellular automaton model for freeway traffic. J. of Physics C 2, 2221.
Nagy, A.M., and Simon, V. (2018). Survey on traffic prediction in smart cities. Pervasive Mobile Comput. 50, 148–163.
Newell, G.F. (1953). A simplified theory of kinematic waves in highway traffic, part I: general theory. Transportation Res. Part B: Methodological 27, 281–287.
Ni, D., and Leonard, J.D. (2005). A simplified kinematic wave model at a merge bottleneck. Appl. Math. Model. 29, 1054–1072.
Niu, K., Zhang, H., Zhou, T., Cheng, C., and Wang, C. (2019). A novel spatio-temporal model for city-scale traffic speed prediction. IEEE Access. https://doi.org/10.1109/ACCESS.2019.2902185.
Pan, Z., Liang, Y., Wang, W., Yu, Y., Zheng, Y., and Zhang, J. (2019). Urban traffic prediction from spatio-temporal data using deep meta learning. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Association for Computing Machinery, New York, NY, USA, pp. 1720–1730.
Park, C., Lee, C., Bahng, H., Tae, Y., Kim, K., Jin, S., Ko, S., and Choo, J. (2020). ST-GRAFT: a novel spatio-temporal graph attention network for accurately forecasting dynamically changing road speed. In Proc.29th ACM Int Conf Inf Knowledge Management, pp. 1215–1224, arXiv:1911.13181.
Park, J., Li, D., Murphy, Y.L., Kristinsson, J.G., Kuang, M.L., and Phillips, A.M. (2014). Intelligent trip modeling for the prediction of an origin–destination traveling speed profile. IEEE Trans. Intell. Transportation Syst. 15, 1039–1053.
Polson, N., and Sokolov, V. (2017). Deep learning for short-term traffic flow prediction. Transportation Res. Part C: Emerging Tech. https://doi.org/10.1016/j.trc.2017.02.024.
Polychronopoulos, A., Tsagkas, M., Amditis, A.J., and Andreone, L. (2007). Sensor fusion for predicting vehicles’ path for collision avoidance systems. IEEE Trans. Intell. Transportation Syst. 8, 549–562. https://doi.org/10.1109/TITS.2007.903439.
Qiu, Y., and Ishaq, S. (2014). A Hidden Markov Model for Short term prediction of traffic conditions on freeways. Transportation Res. Part C: Emerging Tech. 43, 95–111.
Qian, L.F., Feng, A., Yu, N., Xu, W., and Wu, Y. (2020). Vehicular networking-enabled vehicle state prediction via two-level quantized adaptive kalman filtering. IEEE Internet Things J. 7, 7181–7193.
Qu, L., Liu, J., Li, W., Ma, D., and Fan, H. (2021). Features injected recurrent neural networks for short-term traffic speed prediction. Neurocomputing 451, 290–304.
Rapant, L., Slaninová, K., Martinovič, J., and Martinović, T. (2016). Traffic speed prediction using hidden Markov models for Czech republic highways. In Agent and Multi-Agent Systems: Technology and Applications, G. Jezic, Y.H. Chen-Burger, R.J. Howlett, and L.C. Jain, eds. (Springer International Publishing), pp. 187–196.
Raza, A., and Zhong, M. (2017). Hybrid lane-based short-term urban traffic speed forecasting: a genetic approach. In 2017 4th International Conference on Transportation Information and Safety (ICTIS) (IEEE), pp. 271–279. https://doi.org/10.1109/ICTIS.2017.8047776.
Shao, Y., and Sun, Z. (2021). Eco-approach with traffic prediction and experimental validation for connected and autonomous vehicles. IEEE Trans. Intell. Transportation Syst. 22, 1562–1572. https://doi.org/10.1109/TITS.2020.2972198.
Shen, G., Chen, C., Pan, Q., Shen, S., and Liu, Z. (2018a). Research on traffic speed prediction by temporal clustering analysis and convolutional neural network with deformable kernels (may, 2018). IEEE Access 6, 51756–51765.
Shen, G., Yu, K., Zhang, M., and Kong, X. (2021). ST-AFN: a spatial-temporal attention based fusion network for lane-level traffic flow prediction. PeerJ. Computer Sci. 7, e470.
Shen, P., Zhao, Z., Zhan, X., Li, J., and Guo, Q. (2018b). Optimal energy management strategy for a plug-in hybrid electric commercial vehicle based on velocity prediction. Energy 155, 838–852.
Shih, C., Huang, P.W., Yen, E.T., and Tsung, P. (2019). Vehicle speed prediction with RNN and attention model under multiple scenarios. In 2019IEEE Intell. Transportation Syst. Conf. (Itsc). https://doi.org/10.1109/ITSC.2019.8917479.
Wave model. Transportation Res. Rec. 2188, 29–36.

Vlahogianni, E., Karlaftis, M., and Golias, J. (2014). Short-term traffic forecasting: where we are and where we’re going. Transportation Res. Part C: Emerging Tech. 43. https://doi.org/10.1016/j.trc.2014.01.005.

Vogel, K. (2003). A comparison of headway and time to collision as safety indicators. Accid. Anal. Prev. 35, 427–433. https://doi.org/10.1016/S0001-4575(02)00022-2.

Wang, C.C., Thorpe, C., and Suppe, A. (2003). LADAR-based detection and tracking of moving objects from a ground vehicle at high speeds. In IEEE IV2003 Intelligent Vehicles Symposium. Proc. (Cat. No.03TH8883), pp. 416–421. https://doi.org/10.1109/IVS.2003.1212947.

Wang, J., Chen, R., and He, Z. (2019). Traffic speed prediction for urban transportation network: a path based deep learning approach. Transportation Res. Part C: Emerging Tech. 100, 372–385.

Wang, J., Gu, Q., Wu, J., Liu, G., and 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.

Wang, J., and Shi, Q. (2013). Short-term traffic speed forecasting hybrid model based on Chaos-Wavelet Analysis-Support Vector Machine theory. Transportation Res. Part C: Emerging Tech. 27, 219–232.

Wang, X., Chen, C., Min, Y., He, J., Yang, B., and Zhang, Y. (2018). Efficient Metropolitan Traffic Prediction Based on Graph Recurrent Neural Network. Preprint at arXiv, arXiv:1811.00740.

Wang, Y., and Infeld, D. (2018). Markov chain Monte Carlo simulation of electric vehicle use for network infrastructure studies. Int J. Electr. Power Energy Syst. 99, 85–94. https://doi.org/10.1016/j.ijepes.2018.01.008.

Williams, B.M., and Hoe, L.A. (2003). Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: theoretical basis and empirical results. J. Transportation Eng. 129, 664–672.

Wu, Y., Tan, H., Qin, L., Ran, B., and Jiang, Z. (2018). A hybrid deep learning based traffic flow prediction method and its understanding. Transportation Res. Part C: Emerging Tech. 90, 166–180.

Wu, Z., Pan, S., Long, G., Jiang, J., and Zhang, C. (2019). Graph wavelet for deep spatial-temporal graph modeling. Proceedings of the 28th International Joint Conference on Artificial Intelligence, 1907–1913.

Xiang, C., Ding, F., Wang, W., and He, W. (2017). Energy management of a dual-mode power-split hybrid electric vehicle based on velocity prediction and nonlinear model predictive control. Appl. Energy 189, 640–653.

Xie, Q., Guo, T., Chen, Y., Xiao, Y., Wang, X., and Zhao, B.Y. (2019a). How do urban incidents affect traffic speed? A Deep Graph Convolutional Network for Incident-Driven Traffic Speed Prediction. Preprint at arXiv, arXiv:1912.01242.

Xie, S., He, H., and Peng, J. (2017). An energy management strategy based on stochastic model predictive control for plug-in hybrid electric buses. Appl. Energy 196, 279–288. https://doi.org/10.1016/j.apenergy.2016.12.112.

Xie, S., Hu, X., Xin, Z., and Brightton, J. (2019b). Pontryagin’s Minimum Principle based model predictive control of energy management for a plug-in hybrid electric bus. Appl. Energy 236, 893–905. https://doi.org/10.1016/j.apenergy.2018.12.032.

Xie, Y., Xiong, Y., and Zhu, Y. (2020). ISTD-GCN: Iterative Spatial-Temporal Diffusion Graph Convolutional Network for Traffic Speed Forecasting. Preprint at arXiv, arXiv:2008.03970.

Xu, M., Dai, W., Liu, C., Gao, X., Lin, W., Qi, G.J., and Xiong, H. (2021). Spatial-Temporal Transformer Networks for Traffic Flow Forecasting. Preprint at arXiv, arXiv:2001.02908.

Yan, M., Li, M., Hongwen, h., and Peng, J. (2018). Deep learning for vehicle speed prediction. Energy Proc. 152, 618–623. https://doi.org/10.1016/j.egypro.2018.09.220.

Yang, H., Dillon, T.S., and Chen, Y.P. (2017). Optimized structure of the traffic flow forecasting model with a deep learning approach. IEEE Trans. Neural Networks Learn. Syst. 28, 2371–2381. https://doi.org/10.1109/TNNLS.2016.2574840.

Yang, X., Yuan, Y., and Liu, Z. (2020). Short-term traffic speed prediction of urban road with multi-source data. IEEE Access 8, 87541–87551.

Yao, B., Chen, C., Cao, Q., Jin, L., Zhang, M., Zhu, H., and Yu, B. (2017). Short-term traffic speed prediction for an urban corridor. Computer-Aided Civil Infrastructure Eng. 32, 154–169.

Ye, F., Hao, P., Qi, X., Wu, G., Borboonsomsin, K., and Barth, M.J. (2019). Prediction-based eco-approach and departure at signalized intersections with speed forecasting on preceding vehicles. IEEE Trans. Intell. Transportation Syst. 20, 1378–1389.

Ye, J., Zhao, J., Ye, K., and Xu, C. (2020). How to build a graph-based deep learning architecture in traffic domain: a survey. IEEE Trans. Intell. Transportation Syst. 1–21.

Yeon, K., Min, K., Shin, J., Sunwoo, M., and Han, M. (2019). Ego-vehicle speed prediction using a long short-term memory based recurrent neural network. Int. J. Automotive Technology 20, 713–722.

Yin, X., Wu, G., Wei, J., Shen, Y., Qi, H., and Yin, B. (2021a). Deep learning on traffic prediction: methods, analysis and future directions. IEEE Trans. Intell. Transportation Syst. 1–17. https://doi.org/10.1109/TITS.2021.3054840.

Yin, X., Wu, G., Wei, J., Shen, Y., Qi, H., and Yin, B. (2021b). Multi-stage attention spatial-temporal graph networks for traffic prediction. Neurocomputing 428, 42–53.

Yu, B., Li, M., Zhang, J., and Zhu, Z. (2019a). 3D Graph Convolutional Networks with Temporal Graphs: A Spatial Information Free Framework for Traffic Forecasting. Preprint at arXiv, arXiv:1903.00919.
Yu, B., Yin, H., and Zhu, Z. (2018). Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting. Proceedings of the 27th International Joint Conference on Artificial Intelligence (Proceedings of the 27th International Joint Conference on Artificial Intelligence), pp. 3634–3640. https://doi.org/10.24963/ijcai.2018/505.

Yu, D., Liu, C., Wu, Y., Liao, S., Anwar, T., Li, W., and Zhou, C. (2019b). Forecasting short-term traffic speed based on multiple attributes of adjacent roads. Knowledge-Based Syst. 163, 472–484.

Yu, J.J.Q. (2021). Citywide traffic speed prediction: a geometric deep learning approach. Knowledge-Based Syst. 187, 106592. https://doi.org/10.1016/j.knosys.2021.106592.

Yu, J.J.Q., and Gu, J. (2019). Real-time traffic speed estimation with graph convolutional generative autoencoder. IEEE Trans. Intell. Transportation Syst. 20, 3940–3951.

Yu, J.J.Q., Marks, C., and Zhang, S. (2021). Long-term urban traffic speed prediction with deep learning on graphs. IEEE Trans. Intell. Transportation Syst. 1–12.

Yu, R., Li, Y., Shahabi, C., Demiryurek, U., and Liu, Y. (2017). Deep learning: a generic approach for extreme condition traffic forecasting. In Proceedings of the 2017 SIAM International Conference on Data Mining (SDM) (Society for Industrial and Applied Mathematics), pp. 777–785. https://doi.org/10.1137/1.9781611974973.87.

Yuan, H., and Lu, G. (2021). A survey of traffic prediction: from spatio-temporal data to intelligent transportation. Data Sci. Eng.

Yue, Y., Yeh, A.G.O., and Zhuang, Y. (2007). Prediction Time Horizon and Effectiveness of Real-Time Data on Short-Term Traffic Predictability. In 2007 IEEE Intelligent Transportation Systems Conference, pp. 962–967.

Zang, D., Ling, J., Wei, Z., Tang, K., and Cheng, J. (2019). Long-term traffic speed prediction based on multiscale spatio-temporal feature learning network. IEEE Trans. Intell. Transportation Syst. 20, 3700–3709.

Zhang, C., Zhang, S., Yu, J.J., and Yu, S. (2021a). FASTGNN: a topological information protected federated learning approach for traffic speed forecasting. IEEE Trans. Ind. Inform. 1.