

## Research Article

# Short-Term Traffic Prediction of the Urban Road Network based on the Intelligent Transportation System

Shuai Jiang 

Shandong Dingxun Intelligent Transportation Co., Ltd, Jinan, Shandong 250101, China

Correspondence should be addressed to Shuai Jiang; 2016120065@jou.edu.cn

Received 9 March 2022; Accepted 29 April 2022; Published 4 July 2022

Academic Editor: Song Jiang

Copyright © 2022 Shuai Jiang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

As an important part of the intelligent transportation system research, traffic prediction is the premise of realizing traffic guidance and can provide decision-making basis for traveler service and traffic management. To realize the macromanagement of the entire road network, it must be based on the traffic information of all road sections in the road network. In fact, short-term traffic information has certain characteristics such as real-time, high-dimensional, nonlinear, and nonstationary characteristics, but the traffic information of the same road section has stability and regularity in different periods, and the short-term traffic state has a self-similarity. This makes short-term traffic information predictable. The prediction is made by using the information of the road sections with detectors related to it, and the dynamic dissimilarity matrix is introduced to deal with the three parameters of flow, speed, and time occupancy at the same time. The quantitative relationship between the traffic information of the nondetector road segment and the known traffic information of other road segments, so as to realize the prediction of the traffic information of the nondetector road segment and obtain the complete traffic information of the regional road network. In addition, we use the actual data of the local road network in a certain area to verify the feasibility of the method.

## 1. Introduction

With the development of the urban scale, the increase of the total road network and the rapid growth of motor vehicle ownership, traffic problems have become one of the most important issues in urban development and management [1]. With the widening of the backward gap between the limited supply of road capacity and the rapid growth of total traffic demand, the contradiction between the traffic demand and the traffic supply has become increasingly prominent, and the problem of traffic congestion has become one of the issues that urban people are very concerned about [2]. Moreover, with the continuous improvement of the quality of life, people's demand for the safety, speed, and convenience of urban transportation is getting higher and higher [3, 4]. The fast traffic network and convenient traffic information services have become topics of increasing concern and discussion [5]. Not only that, if traffic congestion is not solved well, it will hinder the economic development of a city or region and affect the image of a city or region [6, 7].

The main goal of urban traffic is to be safe, orderly, and smooth. Safety means that road traffic conditions are good, vehicles pass in compliance with safety regulations, and the less traffic accidents and the losses they cause, the better [8]. It is to maximize the road capacity, and the traffic demand and supply are in a balanced and stable state, and the less delays caused by traffic congestion, the better [9]. With the growing demand for traffic management and the increase in the total number of roads and motor vehicle ownership, urban traffic management cannot rely only on traditional management methods and technologies but must develop an Intelligent Transport System (ITS), among which the current development focus should be on an Intelligent Traffic Management System (ITMS), which mainly includes the following systems: adaptive traffic control, traffic flow collection, traffic guidance, road traffic video surveillance, electronic police (including new electronic police systems), traffic incident detection, comprehensive traffic command, and public-oriented traffic information service platform [10–12]. Through the development of intelligent

transportation, it is the only way to solve the problems of traffic congestion and traffic safety, improve the level of urban traffic management, and continuously meet the increasingly high travel requirements of urban people.

Traffic prediction refers to the use of intelligent calculation methods to measure the current or future traffic information of the target area traffic system based on the historical and existing traffic and related factors. It is the premise and foundation of traffic management and traffic travel information services. Reasonable and accurate road traffic state forecasts can actively guide residents' travel and realize traffic guidance, so as to alleviate traffic congestion and improve the efficiency of urban road use [13].

The short-term changes of traffic information are affected by many factors, which are random and uncertain. The problem to be solved in traffic forecasting is how to conduct a systematic analysis from traffic information changes with randomness and uncertainty [14]. Based on traffic information data from various sources, combined with other influencing factors, we find out the regularity and establish the corresponding forecasting methods and models that are used to predict changes in traffic information for several time periods in the future. Therefore, real-time and accurate traffic prediction is the premise and key to realizing traffic guidance and traffic control. The quality of traffic prediction results is directly related to the effect of traffic guidance and control [15].

The data obtained by the detector are counted according to a certain time interval to obtain the required traffic information time series. The shorter the statistical interval, the greater the influence of disturbance on the data sequence, and the stronger the relationship between nonlinearity and uncertainty, resulting in an indistinct relationship between observations [16]. If the statistical interval is too long, the changing trend of traffic information cannot be tracked in real time. By comparing the prediction effects of different time intervals, the results show that when the time interval is 5 minutes, the error tends to be stable and remains at a relatively low level; when the interval is less than 5 minutes, the error is larger. When the interval is greater than 5 minutes, the error is small, but the predicted trend is not obvious [17, 18].

The data of short-term traffic prediction studied in this paper come from the collection of intelligent transportation system equipment (the data in this paper mainly come from the bayonet detection of fixed detectors). Based on these data, the traffic information in the next 5 minutes in the road network is predicted. The predicted future traffic conditions can be released through the intelligent traffic information service system to provide decision-making basis for pedestrians and management departments [19, 20].

## 2. Method and Theory

The road section without detectors refers to the road section in which there is no fixed detector installed to collect traffic information. To predict the traffic parameters of these nondetector road segments by using the data of the detector road segments in the road network, it is necessary to find out

the relationship between the two parameters. Traffic parameters are indicators that reflect traffic status. Traffic volume, density, and speed are the main parameters to describe the basic characteristics of traffic. These three parameters are interrelated and restrict each other. Among them, flow refers to the traffic volume passing through a fixed point within a specified time; density refers to the number of vehicles on a lane of a unit length at a certain time, which changes with time and interval and cannot reflect the relationship between vehicles and the speed, so the occupancy rate is often used to express traffic density. Therefore, flow, speed, and time occupancy are three basic parameters that describe traffic.

In order to achieve this goal, we must first find out the sample points in the road network that have a similar dynamic development trend to the nondetector road section under study. In this paper, cluster analysis and discriminant analysis are used for processing, and the problem is transformed into the road section under study. The quantitative relationship between the parameters of these sample points is determined by using regression analysis, and the premise of regression analysis is that there is a certain correlation between the targets under study. This paper firstly analyzes the correlation of road sections and then obtains regression analysis. The research ideas of this chapter on the parameter prediction of detectorless road sections are shown in Figure 1.

*2.1. Research Object.* In this study, some sections of a main road are selected as the research objects, and each section is divided into two directions according to the driving direction (Figure 2). The roads in Figure 2 are very common in the study area and are representative.

The road section information in the above road network topology structure diagram is studied. There are 10 road sections in the selected road network. The traffic information of each road section is known. The research parameters include flow, speed, and time occupancy. For parameter attributes, the data on April 12, 2021, were selected for research in clustering and discriminant analysis. In order to show the changing trend over a longer period of time, each time period was merged into 15 min traffic information, of which the traffic data of 20 time periods from 8:00 to 11:00 are used as analysis data, and the data of 20 time periods from 11:00 to 16:00 are used as verification data. In order to verify the feasibility of the method proposed in this paper and test the prediction results, the road section 10 is regarded as a nondetector section here, and the other 9 sections are regarded as detector sections.

### 2.2. Research Method

*2.2.1. Clustering Problem Analysis Process.* The cluster analysis in this paper is to find out the statistics that can measure the similarity among the various road sections according to the three observation indicators of traffic, speed, and time occupancy rate of the selected road sections. Then, we use these statistics as the basis for classification and

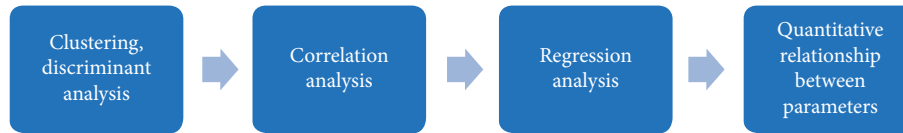


FIGURE 1: The prediction idea of road segment parameters without detectors.

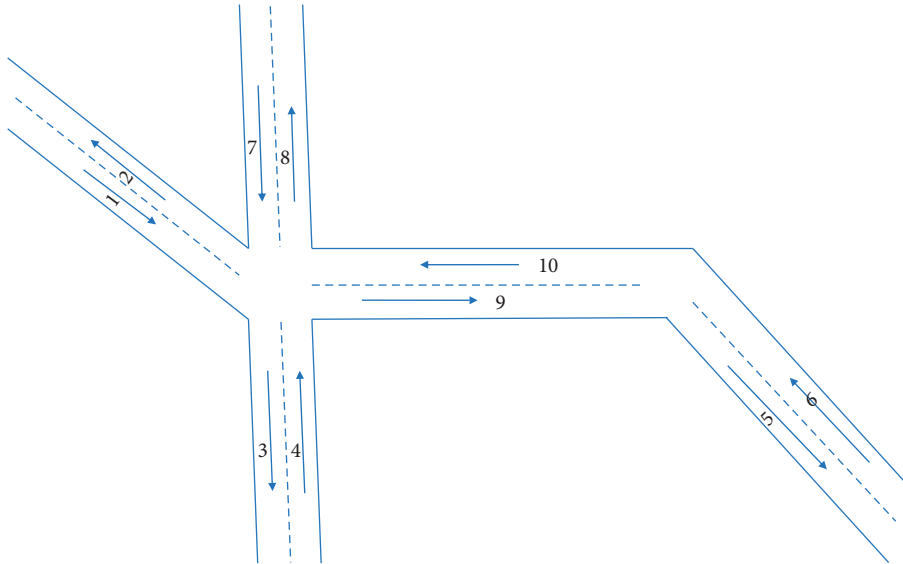


FIGURE 2: The road network topology diagram.

classify the data objects into multiple classes, so that the closely related ones are clustered into a small taxonomic unit, and the distant ones are clustered into a large taxonomic unit, until all samples or indicators are clustered, or the set stopping condition is reached.

(1) *Selecting different road sections in the road network as clustering factors.* First of all, the factors of cluster analysis should meet the requirements of cluster analysis and reflect the purpose of the research. In order to study the short-term traffic prediction of road sections without detectors in the road network, this paper selects the 10 road segments in the above section as clustering factors, and each clustering factor takes into account the three attributes of traffic, speed, and time occupancy. Secondly, the selected variables should not have orders of magnitude difference, otherwise the results will be biased. Data transformations were normalized before analysis. The following table lists the value (unit/vehicle) of the attribute of traffic in 10 road segments. Speed and time occupancy also have a data matrix in the same format as traffic. The analysis in this chapter is based on this matrix-based data processing.

(2) *Data transformation processing.* In the process of cluster analysis, it is necessary to perform some mutual comparison operations on each data set, and the original data of different attributes often affect this comparison operation due to different measurement units. Therefore, it is necessary to transform the original data before the cluster analysis, and we should turn it into a new value according to a certain operation to eliminate the influence of different

measurement units on the value. The data will not affect the dynamic trend of the original data. All data are normalized. By using the min-max normalization method, the original data is linearly transformed, and the original value  $X$  is mapped to a value in the interval  $[-1,1]$  through min-max normalization.

(3) *Define the distance between samples and the distance between classes.* After standardizing the data, cluster analysis can be performed according to the processed data. The clustering is based on the distance between samples. According to different distance calculation standards, there are absolute value distance, Euclidean distance, and Minkow's distance, base distance, Chebyshev distance, etc. When calculating the distance in this article, the absolute value distance is used.

(4) *Computing cluster statistics.* The clustering statistic is new data calculated from the transformed data, which is used to indicate the closeness of the relationship between each road segment. Since this paper studies the three parameters of flow, speed, and time occupancy at the same time, and each parameter considers the value of multiple time periods, in order to ensure that each attribute of the same road segment is clustered into the same class, this paper introduces the dynamic dissimilarity matrix. The dynamic dissimilarity matrix is a method to measure the dynamic dissimilarity between sample points based on stage growth. It considers two more attributes other than the similarity coefficient.

(5) *Choose a clustering method.* In this paper, the system clustering method is selected, and the closely related road

sections are clustered into one category, otherwise they are divided into different categories. Systematic clustering is a method of merging classes one by one. After specifying the distance between samples and the distance between classes,  $n$  samples are first classified into one class; at the beginning, each sample forms its own class, the distance between classes is equal to the distance between samples; then, the two classes with the closest distance are merged; this process is repeated until it stops when the specified number of classes is reached. Finally, this paper is selected to be divided into 2 categories.

**2.2.2. Correlation Relationship.** Correlation analysis is to study whether there is a certain dependency between phenomena and to explore the direction and degree of correlation of the phenomenon with specific dependency. It is a statistical method to study the correlation between random variables. It is worth noting that there is a correlation between things, not necessarily a causal relationship, but may only be an accompanying relationship. But if there is a causal relationship between things, the two must be related. Regression analysis is a mathematical statistical analysis and processing of causal influencing factors and prediction objects. If there is a certain relationship between the independent variable and the dependent variable, then it is meaningful to establish a regression equation. If there is no correlation between the respective variables and the dependent variable in the correlation analysis, there is no need to do regression analysis.

Correlation analysis and regression analysis are two stages of generalized correlation analysis, and they are closely related.

- (1) Correlation analysis is the basis and premise of regression analysis, and regression analysis is the in-depth and continuation of correlation analysis. Correlation analysis needs to rely on regression analysis to show the specific form of quantitative correlation between variables, while regression analysis needs to rely on correlation analysis to show the degree of correlation between quantitative changes and variables. If there is no highly correlated relationship between the variables, it is meaningless to use regression analysis to seek the specific form of the correlation, before making a correct judgment on whether the variables are correlated or the direction.
- (2) Since correlation analysis only studies the direction and degree of correlation between variables, it cannot infer the specific form of the relationship between variables, nor can it infer the change of another variable from the change of one variable. Therefore, in the specific application process, the purpose of research and analysis can only be achieved by combining correlation analysis and regression analysis. Therefore, whether the independent variable factor is related to the predicted object of the dependent variable, what is the degree of their correlation, and the judgment of the possibility of judging the degree of correlation are the problems that must be solved before regression analysis.

**2.2.3. Regression Prediction.** Regression analysis is to determine the causal relationship between variables by specifying dependent variables and independent variables, establishing a regression model, and solving each parameter of the model according to the measured data and then evaluating whether the regression model can fit the measured data well. If the fit is good, further predictions can be made based on the independent variables. Regression can basically be regarded as a fitting process; that is, the most appropriate mathematical equation is used to fit the original observed data consisting of a dependent variable and multiple independent variables. If the relationship between independent variables and dependent variables in regression analysis is linear, it is called linear regression analysis, and this paper selects linear regression analysis.

### 3. Results and Discussion

**3.1. Clustering and Discriminant Analysis Results.** The discriminant criterion used in the cluster analysis and discriminant analysis in this paper are the distance discriminant method, in which the distance is represented by the dynamic dissimilarity. In order to verify the proposed short-term traffic prediction without detectors, road segment 10 is regarded as a road segment without detectors, and the traffic information of the other 9 road segments are known because the basis of cluster analysis and discriminant analysis are all dynamic dissimilarities. The degree matrix and the clustering rules are also the same. Therefore, the same result can be obtained by clustering analysis of 9 road sections with known traffic information first and then carrying out discriminant analysis and directly clustering analysis of 10 road sections. In this paper, 10 road sections are directly clustered at the same time in research and analysis, and the system clustering method is adopted to obtain the clustering result as shown in Figure 3.

In this paper, all road sections are divided into two categories, namely, road sections 5, 6, 3, 1, and 2 are the first category; and road sections 4, 9, 7, 8, and 10 are the second category. It can be seen from the road network topology Figure 3 that the road segments directly related to road segment 10 are 4, 9, 7, 8, and 6 road segments. Since the road network selected in this study is small in scope, the second type after clustering, only some of the 5 road segments are included.

Through the above clustering and discriminant analysis, it is possible to obtain road sections with similar laws to the studied road sections without detectors. In order to predict the traffic information of road sections without detectors, it is necessary to use the clustering results to perform regression analysis to obtain road parameter quantitative relationships between them. In order to avoid "false regression," correlation analysis should be carried out before regression analysis. Regression analysis is only meaningful if there is a correlation between the studied road segments.

**3.2. Regression Prediction Results.** Select the traffic information of sections 4, 8, and 7 as the independent variable, and the traffic information of section 10 as the dependent

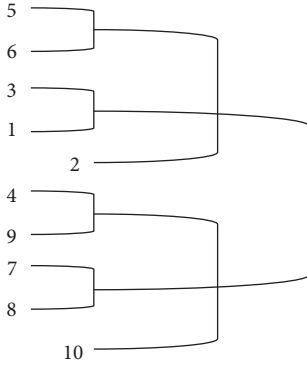


FIGURE 3: Clustering results of road segments.

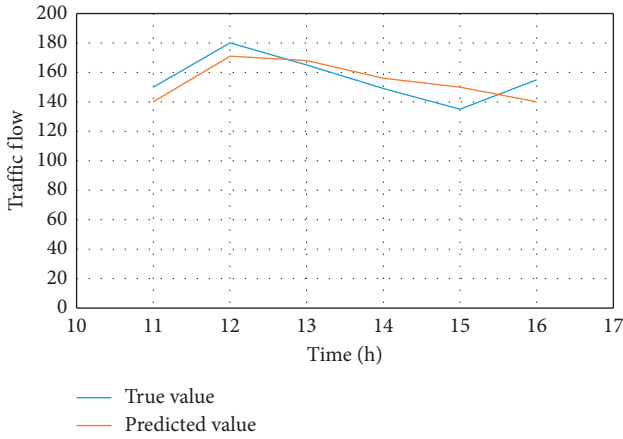


FIGURE 4: Comparison of traffic forecast and its observations.

variable. Regression analysis is performed on the three parameters of flow, speed, and time occupancy with the data before normalization. According to the previous description, it can be known that the three attributes of road section 10 are  $X_{10,1}$ ,  $X_{10,2}$ , and  $X_{10,3}$ , respectively, then the expression of flow

$$X_{10,1} = 145.6234 + 0.15X_{4,1} + 0.231X_{7,1} - 0.415X_{8,1}. \quad (1)$$

Similarly, the expression for speed is

$$X_{10,2} = 11.525 + 0.342X_{4,2} + 0.425X_{7,2} + 0.253X_{8,2}. \quad (2)$$

The time occupancy is expressed as

$$X_{10,3} = 28.436 - 0.534X_{4,3} - 0.063X_{7,3} + 0.298X_{8,3}. \quad (3)$$

**3.3. Analysis of Prediction Results.** In order to further verify the results of the above prediction, the data of 20 time periods from 11:00 to 18:00 were selected as the verification, and the predicted value of road segment 10 are compared with the original observation value of road segment 10. The results are shown in Figures 4–6.

It can be seen from the above three figures that the prediction results of section 10 obtained by the method proposed in this paper have the same trend as the observed values, and the error is quite small.

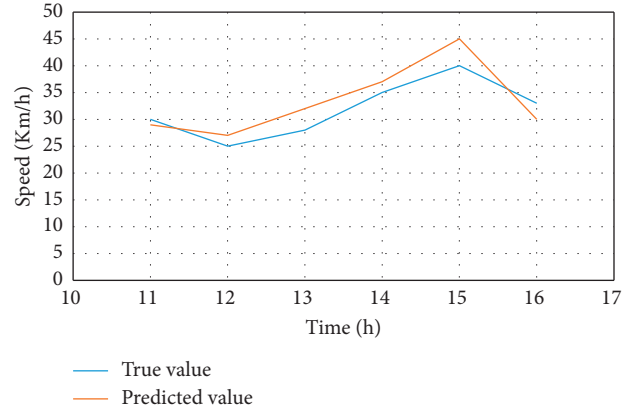


FIGURE 5: Comparison of speed predictions and observations.

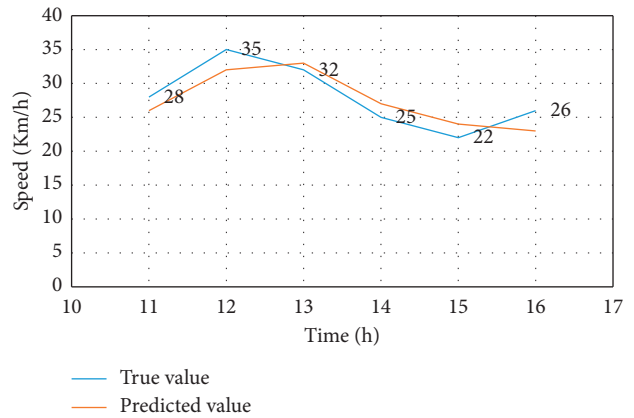


FIGURE 6: Comparison of speed prediction with other road segment observations.

The average error between the predicted value and the actual value in each period in Figure 5 is 3.12%, the average error between the predicted value and the actual value in each period in Figure 6 is 4.68%, and the average error between the predicted value and the actual value in each period is 1.17%.

### 4. Conclusion

- (1) This paper mainly starts from the prediction of traffic information parameters to solve the practical problem of dynamic traffic guidance and control, predicts traffic information parameters of road sections without detectors, and is committed to realizing real-time dynamic traffic information parameter prediction of all road sections in the entire road network. Three categories of traffic information parameters, flow, time occupancy, and average speed, are used to cluster road segments. On this basis, discriminant analysis is used to obtain the attribution class of the studied road segment, and correlation analysis is performed on several road segments in the class.
- (2) This paper selects a part of the regional road network composed of main roads in a certain city for analysis

and verification. Each road segment in the road network has traffic parameter attributes and time series values and then performs cluster analysis based on the three-dimensional array data table. Regression analysis predicts traffic parameters for detectorless road segments, and the analysis is validated with real data. Since only the traffic information parameters of the same road section are used to predict the traffic information of the nondetector road section, the traffic information will be lost, which will affect the prediction result and the uncertainty of the traffic itself. In practice, it is also necessary to periodically correct the model with the manual detection data of the detector-free road section. On the basis of these traffic parameters, it is possible to further predict the travel time, the traffic information status, and the optimal duration of signal lights, providing a basis for the traffic information service system and traffic management system.

- (3) Compared with other models, the results show that the new model in this paper is better. The road conditions of the current model are relatively simple, and the new model will be applied to other more complex roads in the future.

### Data Availability

The figures used to support the findings of this study are included within the article.

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### Acknowledgments

The authors would like to acknowledge the techniques contributed to this research.

### References

- [1] S. Cullinane and K. Cullinane, "Attitudes towards traffic problems and public transport in the Dartmoor and Lake District National Parks," *Journal of Transport Geography*, vol. 7, no. 1, pp. 79–87, 1999.
- [2] R. X. Zhong, A. Sumalee, T. L. Pan, and W. H. Lam, "Optimal and robust strategies for freeway traffic management under demand and supply uncertainties: an overview and general theory," *Transportmetrica: Transportation Science*, vol. 10, no. 10, pp. 849–877, 2014.
- [3] A. Hadayeghi, A. S. Shalaby, and B. N. Persaud, "Safety prediction models: proactive tool for safety evaluation in urban transportation planning applications," *Transportation Research Record*, vol. 2019, no. 1, pp. 225–236, 2007.
- [4] H. Kim, "An analysis in satisfaction of urban tourism influenced by convenience of transportation: Mokpo city in Jeonnam province," *Culinary Science and Hospitality Research*, vol. 25, no. 9, pp. 1–6, 2019.
- [5] J. Zhang, Q. Huang, H. Wu, and Y. Liu, "A shallow network with combined pooling for fast traffic sign recognition," *Information*, vol. 8, no. 2, p. 45, 2017.
- [6] M. E. Porter, "New strategies for inner-city economic development," *Economic Development Quarterly*, vol. 11, no. 1, pp. 11–27, 1997.
- [7] J. L. Nasar, "The evaluative image of the city," *Journal of the American Planning Association*, vol. 56, no. 1, pp. 41–53, 1990.
- [8] N. Buch, S. A. Velastin, and J. Orwell, "A review of computer vision techniques for the analysis of urban traffic," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 3, pp. 920–939, 2011.
- [9] D. J. Graham and S. Glaister, "Road traffic demand elasticity estimates: a review," *Transport Reviews*, vol. 24, no. 3, pp. 261–274, 2004.
- [10] K. Ambak, R. Atiq, and R. Ismail, "Intelligent transport system for motorcycle safety and issues," *European Journal of Scientific Research*, vol. 28, no. 4, pp. 600–611, 2009.
- [11] S. Samadi, A. P. Rad, F. M. Kazemi, and H. Jafarian, "Performance evaluation of intelligent adaptive traffic control systems: a case study," *Journal of Transportation Technologies*, vol. 2, no. 3, p. 248, 2012.
- [12] E. Macioszek and A. Kurek, "Extracting road traffic volume in the city before and during COVID-19 through video remote sensing," *Remote Sensing*, vol. 13, no. 12, p. 2329, 2021.
- [13] W. Min and L. Wynter, "Real-time road traffic prediction with spatio-temporal correlations," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 4, pp. 606–616, 2011.
- [14] S. Ishak and H. Al-Deek, "Performance evaluation of short-term time-series traffic prediction model," *Journal of Transportation Engineering*, vol. 128, no. 6, pp. 490–498, 2002.
- [15] T. Weiskircher, Q. Wang, and B. Ayalew, "Predictive guidance and control framework for (Semi-)Autonomous vehicles in public traffic," *IEEE Transactions on Control Systems Technology*, vol. 25, no. 6, pp. 2034–2046, 2017.
- [16] D. C. Gazis and C. H. Knapp, "On-line estimation of traffic densities from time-series of flow and speed data," *Transportation Science*, vol. 5, no. 3, pp. 283–301, 1971.
- [17] G. Leduc, "Road traffic data: collection methods and applications," *Working Papers on Energy, Transport and Climate Change*, vol. 1, no. 55, pp. 1–55, 2008.
- [18] H. Bar-Gera, "Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: a case study from Israel," *Transportation Research Part C: Emerging Technologies*, vol. 15, no. 6, pp. 380–391, 2007.
- [19] G. Gui, Z. Zhou, J. Wang, F. Liu, and J. Sun, "Machine learning aided air traffic flow analysis based on aviation big data," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 5, pp. 4817–4826, 2020.
- [20] A. Hofleitner, R. Herring, P. Abbeel, and A. Bayen, "Learning the dynamics of arterial traffic from probe data using a dynamic Bayesian network," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 4, pp. 1679–1693, 2012.