



A Novel Traffic Speed Prediction for Road Weaving Sections: Incorporating Traffic Flow Characteristics

Yuming ZHOU¹, Min ZHANG², Chi ZHANG³, Zilong XIE⁴, Bo WANG⁵

Original Scientific Paper
Submitted: 22 Oct. 2023
Accepted: 18 Mar. 2024

¹2023021065@chd.edu.cn, Chang'an University, School of Highway

²Corresponding author, minzhang@chd.edu.cn, Chang'an University, School of Transportation Engineering

³zhangchi@chd.edu.cn, Chang'an University, School of Highway

⁴2022021075@chd.edu.cn, Chang'an University, School of Highway

⁵wb1010110wb@chd.edu.cn, Chang'an University, School of Highway

ABSTRACT

Weaving sections on roads are crucial areas with high concentrations of mandatory lane changes, which can increase the likelihood of traffic accidents. Speed is a key factor in determining traffic safety, and the development of an accurate speed prediction method is essential for improving safety in weaving sections. While current methods are effective in predicting speeds in straightforward situations, they face challenges in more complex scenarios such as weaving sections. This study presents a refined traffic speed prediction approach specifically designed for weaving sections in order to tackle the aforementioned issue. Initially, novel variables were formulated to capture the unique traffic flow attributes present in weaving sections, which distinguish them from standard road segments. Subsequently, supplementary empirical variables that are known to impact speed were incorporated. We conducted a variable importance assessment to ascertain the extent and direction of each variable's contribution. Lastly, variables with significant positive effects were chosen as inputs for three machine learning algorithms: Random Forest (RF), Backpropagation Neural Network optimised with Genetic Algorithm (BPNN-GA) and Support Vector Regression (SVR). This method was evaluated using aerial footage from five distinct weaving sections in China, maintaining an approximately 3 km/h prediction error. In addition, the study also finds that the speed distribution in weaving sections is negatively correlated with the number of lane-changing. Vehicles experience deceleration at both on-ramp and off-ramp, with a more significant deceleration occurring at the on-ramp. Speed is significantly affected by short length, number of lanes and proportion of large vehicles. The proposed method can be embedded into intelligent traffic systems for safe speeds of autonomous vehicles in weaving sections. Reconstructing spatiotemporal patterns of traffic congestion, predicting traffic accidents and implementing active traffic management strategies in weaving sections could be investigated in the future.

KEYWORDS

traffic safety; weaving sections; speed prediction method; vehicle trajectory data; variable importance.

1. INTRODUCTION

The road infrastructure throughout the world continues to expand year by year. By the end of 2022, the total road mileage in the United States had accumulated 6.6 million kilometres, while China's road network had surpassed 5.5 million kilometres. Alongside this expansion, there has been a notable increase in the interchange construction at the crucial intersections of road infrastructures, which serve a vital role in segregating and directing traffic flows. In recent years, cloverleaf configurations have gained increasing popularity in these interchanges, primarily due to land use constraints. This research focuses on weaving sections



This work is licensed under a Creative Commons Attribution 4.0 International License.

Publisher:
Faculty of Transport and Traffic Sciences,
University of Zagreb

within these cloverleaf interchanges, as they are anticipated to contribute significantly to traffic congestion and traffic accidents [1, 2]. The diminished safety level within these weaving sections can be attributed to the following three key characteristics: first, there is a spatial concentration of mandatory lane-changing behaviours; second, there is a coexistence of both large and small vehicles; and third, there is a higher traffic speed dispersion.

Improving traffic safety at weaving sections is of crucial significance and can be achieved through the lenses of road design and operational management. Researchers have examined traffic collision and traffic conflict models to pinpoint design shortcomings in weaving sections, ultimately aiming to refine the design process and enhance safety and reliability [3]. Nonetheless, the majority of the road network has already been established, limiting the scope for safety enhancements through design improvements. On the other hand, in the field of operation management, Intelligent Transport System (ITS) is a widely employed tool for executing various functions, such as real-time traffic monitoring, traffic flow prediction, travel time prediction and route planning [4, 5]. Within this context, traffic speed plays a crucial role as a fundamental input for these functions, and its accuracy directly influences the effectiveness of ITS in implementing operation and management strategies. Accurate speed prediction can aid in developing a more appropriate merging speed management strategy for vehicles entering weaving sections. This can help reduce the rate of traffic accidents in these sections by matching the speed of the merging traffic flow with that of inside the sections. Additionally, it can provide information on the level of service and congestion and assist in identifying congestion, providing path-planning services for connected vehicles in advance. This also can aid the traffic department in developing more reasonable traffic control strategies. In the future, the speed indicator will be a crucial factor in ITS for mixed traffic flow of connected and autonomous vehicles. Its accuracy will directly impact the rationality of implementing other strategies in the ITS system. Therefore, enhancing the accuracy of traffic speed estimation is beneficial in elevating the safety level of road weaving sections from operational management.

Many methods have been employed for predicting traffic speeds. The most classical approach is the Lighthill-Whitham-Richards (LWR) model, which necessitates various types of data as inputs, including traffic speed, traffic density and traffic flow data [6]. In recent times, there has been a growing interest in methods that rely merely on speed data, categorising them into statistical technique methods, data-driven methods and hybrid methods [7, 8]. The first category contains methods that utilise small sample data and traffic flow theory to investigate the pattern of traffic speed change. Notable techniques in this category include the Autoregressive Integrated Moving Average (ARIMA) [9, 10] and the Hidden Markov Model (HMM) [11]. The second category is data-driven methods which primarily include machine learning and deep learning methods, such as Support Vector Machines (SVM) [12], Recurrent Neural Networks (RNN) [13], Convolutional Neural Networks (CNN) [14] and others. Hybrid approaches often combine data-driven methods with theoretical traffic flow models. Recent approaches include Physics-Informed Neural Networks (PINN) [15] and a hybrid method that combines Stacked Autoencoder (SAE) with Long Short-Term Memory (LSTM) [16].

However, these methods concentrate on data fitting and often lack a deep exploration of underlying causes of speed changes, resulting in models with limited interpretability. Additionally, it should be noted that these methods are trained on extensive traffic data with spatial variations, theoretically making them effectively applicable to any road location. Nevertheless, in intricate and complex traffic environments, especially in weaving sections, these methods often exhibit limited prediction accuracy [17, 18]. To address such limitations, this paper proposes a novel method that leverages the traffic flow characteristics of the weaving sections to address the aforementioned challenges. This method has the following key features:

- 1) Refined and targeted: Previous studies have often overlooked the influence of spatial heterogeneity data on prediction outcomes. As a result, the resultant “generic” prediction methods for traffic speeds are broadly applicable across entire roads but tend to exhibit reduced accuracy in specific intricate spatial locations [19, 20]. However, these intricate locations play a particularly crucial role in improving traffic

safety. To address this, we refine the generalised speed prediction model and construct a specialised model tailored for weaving sections to improve prediction accuracy in these weaving sections.

- 2) **Enhanced interpretability:** Previous studies on speed prediction have often relied on data-driven methods such as machine learning and deep learning, to analyse speed change patterns and trends, with a primary focus on accurately fitting the data rather than uncovering the underlying causes of speed variation patterns. In this study, we take a novel approach by drawing from traffic flow theory and the actual speed distribution within weaving sections. We have introduced three distinct types of input variables designed to capture the essence of traffic flow in weaving sections. Furthermore, our research demonstrates the extent to which these innovative variables contribute to variations in traffic speed. This method allows us to explain the main factors driving changes in traffic speed across various weaving section configurations.
- 3) **Reliability and wide applicability:** We validated the reliability of our method by utilising real aerial video data. It is worth noting that many prior studies have relied on the Next Generation Simulation (NGSIM) dataset, collected by the Federal Highway Administration (FHWA) of the United States. Nevertheless, this dataset may not accurately reflect the conditions on China roads, given disparities in route design standards, speed limits and driving rules. In our study, we conducted fieldwork to capture aerial videos from four weaving sections situated in Xi'an, China. We then extracted vehicle trajectory data for model validation. Furthermore, we compared our approach with the classical Highway Capacity Manual 6th (HCM6) method and the state-of-the-art method, to prove the reliability and applicability of our method for weaving sections in western China.

The paper is structured as follows. Section 2 summarises related work and highlights the contributions of this paper. Section 3 details the proposed traffic speed prediction method for weaving sections. Section 4 validates the effectiveness of the proposed method by comparing predicted speeds from the following methods: the proposed method in this paper, the classical HCM6 method and the state-of-the-art LSTM, using collected real speed data as a benchmark. Section 5 discusses the findings and offers future research suggestions.

2. LITERATURE REVIEW

This section reviews previous traffic speed prediction methods, categorised into three groups: (i) statistical technique methods, (ii) data-driven methods, and (iii) hybrid methods.

Statistical technique methods, such as ARIMA [9, 10], HMM [11] and Kalman Filter [21, 22], are primarily suited for scenarios with limited data and simple traffic conditions [23]. For example, ARIMA makes the unrealistic assumption that the traffic states follow a stable process with constant mean, variance and autocorrelation values, which does not align with real-world complexities. In addition, the HCM6 method constructs a regression model to achieve traffic speed prediction on weaving sections, but its accuracy has been demonstrated to be insufficient [24]. Given this, several studies have attempted to enhance its predictive performance. Xu et al. [25] proposed a new speed prediction method that converges to the basic road section under low weaving flow or substantial weaving lengths, surpassing the HCM6 method. Sun et al. [26] proposed a segmented and modified model for speed prediction in weaving sections, employing a genetic algorithm for parameter calibration. The results showed a total error of less than 10%, with an average prediction error of 9.52% for weaving traffic speed and 6.64% for non-weaving traffic speed. However, it is worth noting that the predictive accuracy of statistical technique methods is relatively low since they struggle to capture non-linear variations in traffic speed.

Data-driven methods, including machine learning and deep learning, have become prevalent in estimating traffic speed, thanks to advancements in Artificial Intelligence (AI) and traffic data science. These methods are well-suited for the wealth of traffic data available and excel at capturing non-linear relationships, effectively addressing the limitations of statistical technique methods. Key data-driven methods for traffic speed prediction encompass LSTM, CNN and SVM, among others. LSTM has gained popularity for its ability to accurately capture non-linear traffic states and forecast speed based on time series [27, 28]. However, it focuses mainly on the temporal dimension of data while ignoring the spatial dimension, which

can be taken into account by CNN. CNN has become a common method for predicting the traffic speed in road networks. Using a two-dimensional spatio-temporal matrix, it transforms traffic states into images and leverages powerful spatial feature extraction capability to predict traffic speeds [29, 30]. Ma et al. [31] compared this method against three deep learning methods – SAE, RNN and LSTM – as well as four widely used methods: Ordinary Least Squares (OLS), K-Nearest Neighbor (KNN), Artificial Neural Network (ANN) and RF. Their model outperforms the others, improving the average accuracy by 42.91% while maintaining reasonable execution time. Wang et al. [32] designed a recursive convolutional neural network structure with error feedback, considering the influence of the nearby road sections and uncovering the implicit linkages between road sections. Ke et al. [33] proposed a two-stream multi-channel convolutional neural network (TM-CNN) to account for the influence of traffic flow on speed.

Several studies have shown hybrid methods can improve prediction accuracy. Hybrid methods encompass a combination of statistical techniques with data-driven methods and a fusion of two data-driven methods [8, 34]. For instance, Tan et al. [35] combined ARIMA with ANN for short-term traffic speed prediction, taking advantage of ARIMA's robust linear fitting ability and ANN's powerful nonlinear mapping ability. Tian et al. [16] combined SAE and LSTM for traffic flow prediction, using SAE to capture spatial features and LSTM to extract temporal features of traffic flow. Hybrid methods may demonstrate better prediction accuracy and stability compared to individual methods [13, 36].

However, data-driven and hybrid methods prioritise data fitting at the expense of understanding the true causes behind changes in traffic speeds, resulting in low method interpretability. These methods are typically generic, designed for basic road segments [17, 18] and road networks [19, 20], using spatio-temporally mixed traffic data (usually NGSIM dataset) without considering the data spatial heterogeneity. Consequently, they tend to perform poorly in specific spatial locations like weaving sections, which have complex traffic flow characteristics. Furthermore, few studies have examined traffic speed prediction in road weaving sections based on data-driven methods.

To address these gaps, this paper proposes a novel traffic speed estimation method tailored for weaving sections. Our work contributes in the following ways. First, a refined model specially designed for speed estimation in road weaving sections is developed. Second, the spatial speed distribution in the weaving sections at various geographic locations is analysed. Three types of variables are innovatively designed to capture the uniqueness of weaving sections compared to regular road sections. They are designed from the perspectives of lane-changing behaviour, distance of vehicles from the weaving section's on/off-ramps and spatial density distribution of vehicles. The individual contributions of these variables to the estimation speed are quantitatively assessed. Additionally, RF, BPNN-GA and SVR are initially considered as potential main algorithms for this paper's method. Third, this study collected aerial videos of weaving sections in various Chinese locations, amassing approximately 600,000 high-precision vehicle trajectory data points for method validation. Finally, method performance is compared with LSTM, a common choice in the recent speed prediction research, and the traditional HCM6 method. The results show our method is superior to other methods, which effectively adapts to various weaving section configurations and geographical locations to provide insightful explanations for speed variations.

3. SPEED PREDICTION METHOD FOR WEAVING SECTIONS

The method comprises the following main steps: designing three types of unique variables to characterise weaving section traffic patterns, creating a database of speed-influencing variables, assessing their significance, and selecting a preferred speed prediction regression algorithm. Refer to *Figure 1* for a visual representation of the procedure.

1) Designing three types of unique variables for weaving sections

These variables are designed from the following perspectives: lane-changing manoeuvre, vehicle and on/off-ramps positional relativity and vehicle space distribution. We consider four headways that describe lane-changing manoeuvres between the lane-changing vehicles and adjacent front-rear vehicles before

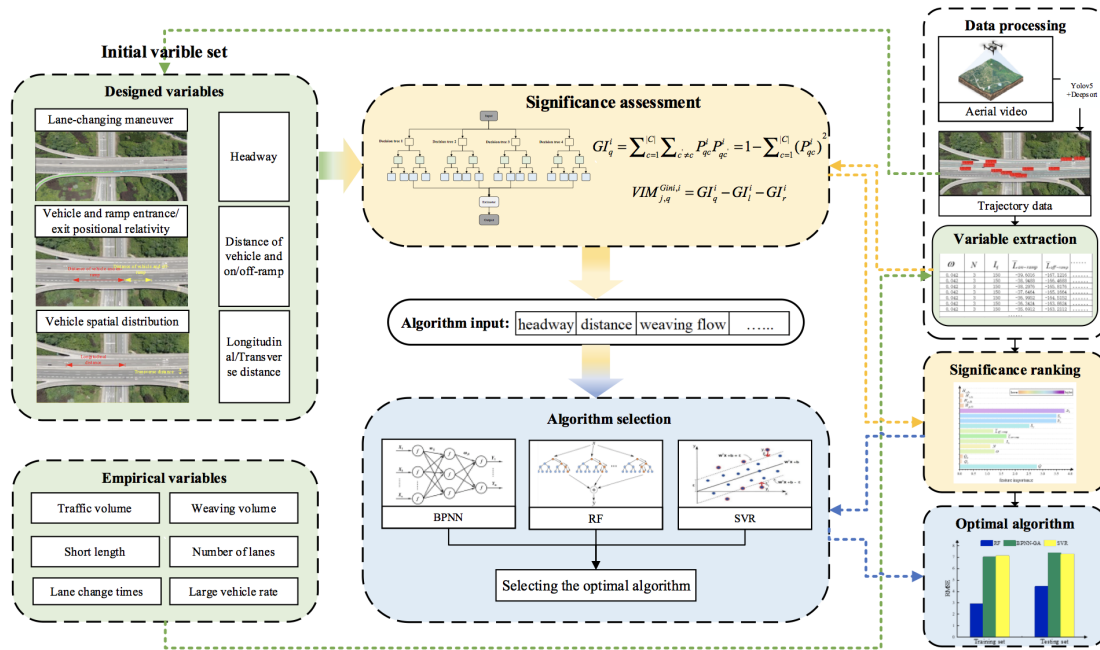


Figure 1 – Method procedure chart

and after the lane changes. We analyse the distance between vehicles and the weaving section on/off-ramps, indicating their positional relativity. We assess longitudinal and transversal vehicle spacings and characterise spatial distribution using a mean and standard deviation of distances between vehicles along a given direction.

2) Building a speed-influencing variables database

The database for speed-influencing variables in weaving sections is enriched with broader empirical variables, in addition to the three types of uniquely designed variables. These empirical variables, including traffic volumes, weaving volumes, number of lane changes, large vehicle proportions and short length (the distance in feet between the endpoints of any barrier markings (solid white lines) that prohibit or discourage lane changing), are sourced from the classical HCM6 method.

3) Assessing variable significance

To streamline method complexity and enhance operational efficiency, we assess variable significance using the RF algorithm, identifying and retaining those variables in the speed-influencing variables database that contribute most significantly to speed variation.

4) Selecting the optimal algorithm

Three prediction algorithms have been chosen: RF, BPNN-GA and SVR. The best-performing algorithm among them will be used as part of this method. Different input data may be suitable for the different algorithms, so choosing multiple algorithms increases method prediction accuracy. The problem of method running time is also considered. More algorithms were not selected in order to avoid long running time, which affects the real-time nature of the result output. Therefore, only three algorithms were selected.

3.1 Design of three types of unique variables

This research explores the lane-changing manoeuvre, vehicle and on/off-ramp positional relativity, and vehicle space distribution in weaving sections to describe complex traffic features of weaving sections. The reasons for this exploration are as follows.

Based on professional theoretical knowledge and extensive video observations of weaving sections, we have formulated a hypothesis that the lane-changing behaviour within these sections may influence the variation of spatial average speed. To test this hypothesis, this paper collected data on spatial average speed and frequency of vehicle lane changes within the weaving sections, as illustrated in Figure 2. Furthermore,

this study extends the analysis to encompass three other geographically weaving sections in addition to the one depicted in *Figure 2*. The speed and lane-changing distributions within these sections exhibit consistency with the patterns illustrated in *Figure 2*.

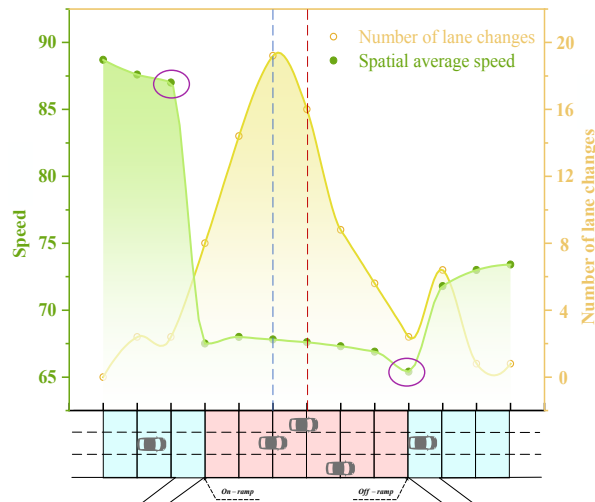


Figure 2 – Relationship between lane-changing frequency and spatial average speed

In this study, we define the “on-ramp” and “off-ramp” as the longitudinal spatial reference points for each weaving section, dividing the road between a pair of reference points into six segments. Each segment is characterised as a spatial unit, with its length stored as a spatial unit length. Beyond these two reference points, additional segments are created on both sides, each delineated by the spatial unit length. Both average spatial speed and number of lane changes shown in *Figure 2* align with the spatial units along their longitudinal positions. From the graphical representation, we make an initial qualitative assessment that the spatial distribution pattern of lane changes in a weaving section is nearly opposite to the pattern of spatial speed distribution. The number of lane changes in the first half of the section exceeds that in the second half, with a peak occurring near the first third of the on-ramp. Therefore, the first variable type introduced is termed “Lane-Changing Manoeuvre Variables”.

Figure 2 clearly illustrates vehicles that exhibit deceleration behaviour at both on-ramp and off-ramp locations of a weaving section, with more significant deceleration at on-ramp than off-ramp. This observation implies vehicles exhibit specific acceleration and deceleration patterns as they approach these ramp points. Consequently, the second variable type introduced is termed “Vehicles and On/Off-Ramps Positional Relativity Variables”.

The longitudinal cross-section speeds exhibit a distinctive pattern characterised as “high-low-high”. Previous research has demonstrated significant variations in the speeds of multi-lane vehicles across the road cross-section [37]. This finding suggests a correlation between speed values and their spatial positions within a weaving section. Consequently, “Vehicle Spatial Distribution Variables” is regarded as the third variable type.

Design of Lane-Changing Manoeuvre Variables

In this paper, headways are employed to illustrate the characteristics of lane-changing manoeuvres. The reason we chose headway over the gap is that we have been experimenting specifically with lane-changing behaviour, and we tried a variety of performances and found headway to work better. Additionally, headway contains more information than the gap. It contains information about the gap as well as the speed of rear cars. These headways are associated with up to five vehicles: the lane-changing vehicle itself, two vehicles directly ahead and behind the lane-changing vehicle in the current lane, and two vehicles in the adjacent lane, one ahead and one behind, after the lane-changing manoeuvre. For each lane-changing manoeuvre, there are up to four headways, representing the distances between the lane-changing vehicle and the other four vehicles mentioned. Theoretically, the headways should be determined when there is no front or rear

vehicle by considering the weaving section’s maximum length and minimum speed. However, for this study, a zero value is assigned to these cases to assess the impact of this variable.

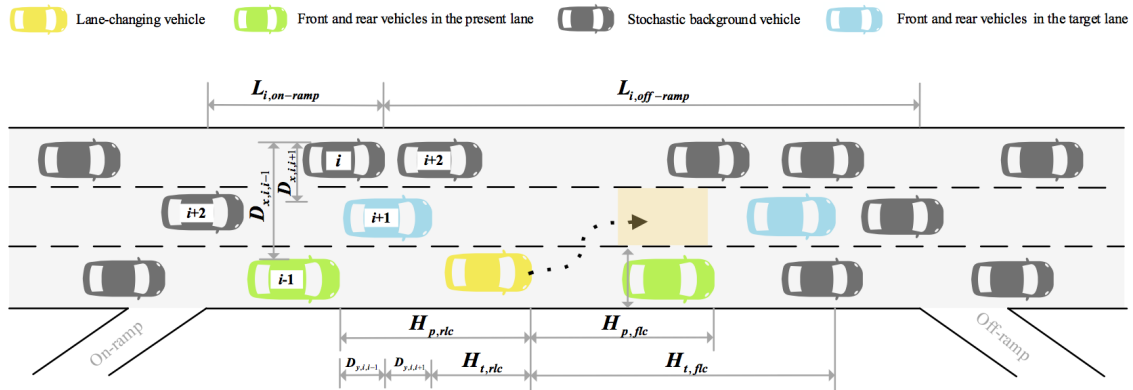


Figure 3 – Diagram of three categories of design variables

Traffic risk increases as headway between vehicles decreases. However, lane-changing manoeuvres commence with the emergence of a need to change lanes and conclude with the selection of an appropriate gap for the lane change. These four headways effectively characterise the lane-changing risk for vehicles within weaving sections, describing the ‘vehicle Lane-Changing Manoeuvre.’ Consequently, the first variable type is represented by these headways, as depicted in Figure 3. The associated equations are as follows:

$$H_{p,rlc} = \frac{L_{p,rlc}}{V_{p,r}} \tag{1}$$

$$H_{p,flc} = \frac{L_{p,flc}}{V_{p,lc}} \tag{2}$$

$$H_{t,rlc} = \frac{L_{t,rlc}}{V_{t,r}} \tag{3}$$

$$H_{t,flc} = \frac{L_{t,flc}}{V_{t,lc}} \tag{4}$$

where $H_{i,j}$ refers to the headways of the j vehicles in the i lane. $i=p$ refers to the present lane, $i=t$ refers to the target lane, $j=flc$ refers to the front vehicle and the lane-changing vehicle, $j=rlc$ refers to the rear vehicle and the lane-changing vehicle, $L_{i,j}$ refers to the distance between j vehicles in the i lane, $V_{i,j}$ refers to the speed between j vehicle in the i lane.

Design of Vehicles and On/Off-Ramps Positional Relativity Variables

As vehicles may accelerate and decelerate when approaching on/off-ramps, vehicle speeds are correlated with the relative positions of vehicles and on/off-ramps locations. In this paper, the average distance between all vehicles and on-ramps signifies the association between vehicles and on-ramps, while the average distance between all vehicles and off-ramps characterises their connection with off-ramps. Figure 3 illustrates the conceptualisation of this variable.

$$\bar{L}_{on-ramp} = \frac{\sum_{i=1}^n L_{i,on-ramp}}{n} \tag{5}$$

$$\bar{L}_{off-ramp} = \frac{\sum_{i=1}^n L_{i,off-ramp}}{n} \tag{6}$$

where $\bar{L}_{on-ramp}$ refers to the mean distance between all vehicles and the small nose points of the solid white line on the on-ramps during the same time interval, $L_{i,on-ramp}$ refers to the distance between the i vehicle and the small nose point of the solid white line on the on-ramps, $L_{i,off-ramp}$ refers to the distance between the i vehicle and the small nose point of the solid white line on the off-ramps and n refers to the number of vehicles during the same time interval.

Design of Vehicle Spatial Distribution Variables

The spatial distribution of vehicles during the same time interval also impacts traffic speed. Two variables represent the spatial distribution of vehicles in both the longitudinal (y) and lateral (x) directions, as depicted in Figure 3. \bar{D}_y denotes the mean distance of all vehicles along the longitudinal road (y) during the same time interval. S_y represents the degree of dispersion in the distribution of all vehicles along the y direction during the same time interval. A larger value \bar{D}_y indicates greater dispersion of vehicles in the y direction, while a higher S_y value signifies more disorganised vehicle distribution in the y direction. Similarly, transverse distribution is represented by \bar{D}_x and S_x . In Figures 4a and 4b, we observe uniform and uneven longitudinal distributions, respectively. Figures 4c and 4d depict situations of uniform and uneven transverse distribution, respectively.

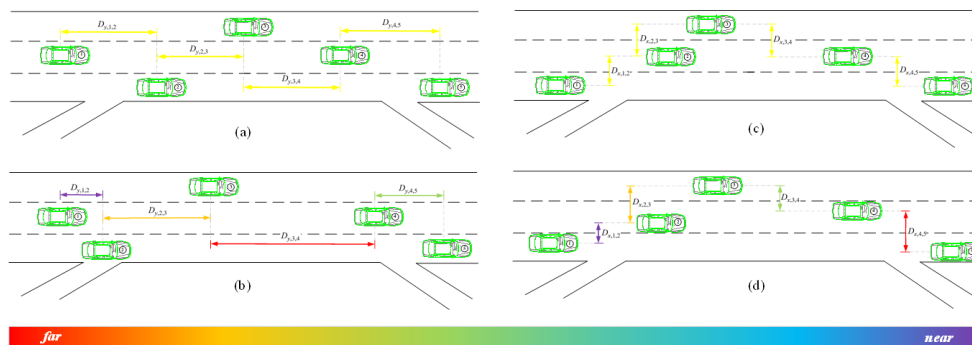


Figure 4 – Schematic diagram of the disorder degree of vehicles in longitudinal and transverse

$$\bar{D}_y = \frac{\sum_{i=1}^n D_{y,i,i+1}}{n - 1} \tag{7}$$

$$S_y = \frac{\sqrt{2 \sum_{i=1}^n (D_{y,i,i+1} - \bar{D}_y)^2}}{n - 1} \tag{8}$$

$$\bar{D}_x = \frac{\sum_{j=1}^n D_{y,j,j+1}}{n - 1} \tag{9}$$

$$S_x = \frac{\sqrt{2 \sum_{j=1}^n (D_{y,j,j+1} - \bar{D}_x)^2}}{n - 1} \tag{10}$$

where \bar{D}_y represents the mean distance between two adjacent vehicles in the front and rear positions after all vehicles are longitudinally arranged during the same time interval, S_y is the standard deviation of the distance between two adjacent vehicles in the front and rear positions after all vehicles are longitudinally arranged in the same time interval. Similarly, \bar{D}_x denotes the mean distance between left and right adjacent vehicles after all vehicles are transversely arranged during the same time interval, and S_x stands for the standard deviation of the distance between left and right adjacent vehicles after all vehicles are transversely arranged in the same time interval.

3.2 Construction of a speed-influencing variable database

The speed-influencing variable pool consists of both empirical variables and design variables. Empirical variables are derived from HCM6’s Weaving Section Model. In HCM6, traffic speeds within the weaving section encompass weaving speeds and non-weaving speeds. Weaving speeds are calculated based on two expected speed limits, short lengths and lane change rates. Among these factors, the lane change rate is connected to the minimum lane change rate, short length and the number of lanes. The minimum lane change rate is correlated with the weaving flow and its corresponding minimum lane change time. On the other hand, the minimum lane change rate, free-flow speed, total flow and the number of lanes affect the speed of non-interleaving lanes. Therefore,

the empirical variables are determined to be traffic volume, weaving volume, lane change time, number of lanes and short length of the weaving section. Furthermore, a component accounting for large vehicle rates has been introduced to consider the impact of large vehicles on speed. The complete set of variables is presented in Table 1.

Table 1 – Description of empirical variables

Type	Variable	Symbol	Explanation
Empirical variables	Traffic volume	Q	Number of traffic entities passing through a section in weaving sections within the corresponding time interval.
	Weaving volume	Q_w	Traffic flow from mainline to ramp or ramp to mainline within the corresponding time interval.
	Lane change occurrence	Q_{lc}	Number of lane-changing occurrences in the entire weaving section within the corresponding time interval.
	Large vehicle rate	ω	Large vehicle proportion in the total traffic volume within the corresponding time interval.
	Number of lanes	N	Number of lanes in the weaving section.
	Short length	L_s	Distance between the solid white lines at the end of the weaving section that prohibits lane-changing.
Designed variables	Lane-changing manoeuvre	$H_{t,fc}$	Headway of the front vehicle and the lane-changing vehicle in the target lane.
		$H_{t,rlc}$	Headway of the rear vehicle and the lane-changing vehicle in the target lane.
		$H_{p,flc}$	Headway of the front vehicle and the lane-changing vehicle in the present lane.
		$H_{p,rlc}$	Headway of the rear vehicle and the lane-changing vehicle in the present lane.
	Vehicles and on/off-ramps positional relativity	$\bar{L}_{on-ramp}$	The average distance between all vehicles and the small nose points of the solid white lines on the on-ramps during the same time interval.
		$L_{off-ramp}$	The average distance between all vehicles and the small nose points of the solid white lines on the off-ramps during the same time interval.
	Vehicle spatial distribution	\bar{D}_x	The average lateral (x-axis) gaps between adjacent vehicles during the same time interval.
		\bar{D}_y	The average longitudinal (y-axis) gaps between adjacent vehicles during the same time interval.
		S_x	The standard deviation of the lateral (x-axis) gaps between adjacent vehicles during the same time interval.
		S_y	The standard deviation of the longitudinal (y-axis) gaps between adjacent vehicles during the same time interval.

3.3 Evaluation variable importance

Although a dataset frequently contains hundreds or thousands of features, choosing the ones to include in the model that substantially impacts the outcome can reduce model complexity while retaining high prediction accuracy. Typical techniques include RF, Principal Component Analysis, etc. This study uses RF to rank the variable significance.

VIM is the variable importance score and GI is the Gini coefficient. Suppose there are j features: $X_1, X_2, X_3, \dots, X_j$ I decision trees and C categories. Now to calculate the Gini coefficient score VIM_j^{Gini} of each feature X_j , namely, the average change of node split impurity of the j feature in all random forest decision trees.

The formula to calculate the Gini coefficient of node q of the i tree is as follows:

$$GI_q^i = \sum_{c=1}^{|C|} \sum_{c' \neq c} P_{qc}^i P_{qc'}^i = 1 - \sum_{c=1}^{|C|} P_{qc}^i{}^2 \tag{11}$$

where C represents categories, P_{qc} represents the proportion of c categories in node q . Intuitively speaking, it represents the likelihood of selecting two random samples from node q , where their category labels are different. The importance of the X_j feature in the tree i node q , namely, the Gini coefficient change before and after the branch of node q is as follows:

$$VIM_{j,q}^{Gini,i} = GI_q^i - GI_l^i - GI_r^i \tag{12}$$

where GI_l^i and GI_r^i respectively represent the *Gini* coefficient of the two new nodes after branching. If the node where the feature X_j appears in the decision tree i is set to Q , then the X_j importance in the i tree is as follows:

$$VIM_j^{Gini,i} = \sum_{q \in Q} VIM_{j,q}^{Gini,i} \tag{13}$$

Assuming that there are I trees in a random forest, then,

$$VIM_j^{Gini} = \sum_{i=1}^I VIM_j^{Gini,i} \tag{14}$$

Finally, the feature importance scores of all the obtained trees are normalised.

$$VIM_j^{Gini} = \frac{VIM_j^{Gini,i}}{\sum_{j'=1}^J VIM_{j'}^{Gini,i}} \tag{15}$$

3.4 Algorithm preference for speed prediction

As the RF name implies, the random forest grows a forest randomly. The forest comprises multiple decision trees, each with no correlation. Regression can be performed using the RF algorithm, which logically meets the research goal. All samples are at the decision tree’s root in its initial state since it has not yet grown. This tree’s sum of residual squares equals the sum of residual squares of the regression. The algorithm’s core principle is that choosing a variable will reduce the sum of squares of the two portions. The law is recirculated at the next bifurcation node. The next categorisation characteristic variable is chosen until a full tree is produced. The weighted mean of the target variable at the leaf node is the expected value of the random forest regression tree.

The RF algorithm offers the following benefits: it introduces unpredictability based on the decision tree, making it less likely to become overfitted. It uses an unbiased estimate for generalisation errors when constructing a random forest. Mutual interactions between features can be discovered during the training process and creating a parallelisation mechanism is relatively straightforward.

4. RESULTS AND DISCUSSION

4.1 Data collection in the weaving sections

Five common A-type weaving sections around Xi’an, China, were selected for data collection. These sections exhibit high weaving volumes, weaving ratios and frequent lane-changing activities. The traffic flow weaving phenomenon is significant, allowing for the extraction of abundant weaving vehicle data to study weaving section characteristics. Basic information about five A-type weaving sections is shown in *Table 2*. An unmanned aerial vehicle (UAV) hovered about 120 meters above these weaving sections, capturing aerial videos during a weekday in July 2022, characterised by cloudy weather and a temperature of 29 degrees.

Table 2 – Basic description of the studied weaving sections

Name	Structure	Short length [m]	Lane	Weaving flow ratio	Weaving ratio	Number of lane-changing [times/h]
HCZ	A	150	3	0.649	0.50	972
SQ(E-W)	A	168	4	0.433	0.56	1530
SQ(W-E)	A	160	4	0.558	0.36	2718
ZB	A	180	4	0.710	0.56	1158
MEL	A	109	5	0.208	0.80	732

4.2 Extraction of highly accurate trajectory data

Processing aerial video into vehicle trajectory data is required before obtaining a speed-influencing variable database from actual weaving sections. In this study, we extracted high-precision vehicle trajectories from weaving sections using Data From Sky platform and Data From Sky Viewer software. A total of 610,200 pieces of

high-precision vehicle trajectories were obtained. The trajectories have a 0.1px spatial accuracy and a 0.0001s time accuracy. Figure 5 depicts the vehicle trajectory extraction interface. Table 3 displays a sample of trajectories.

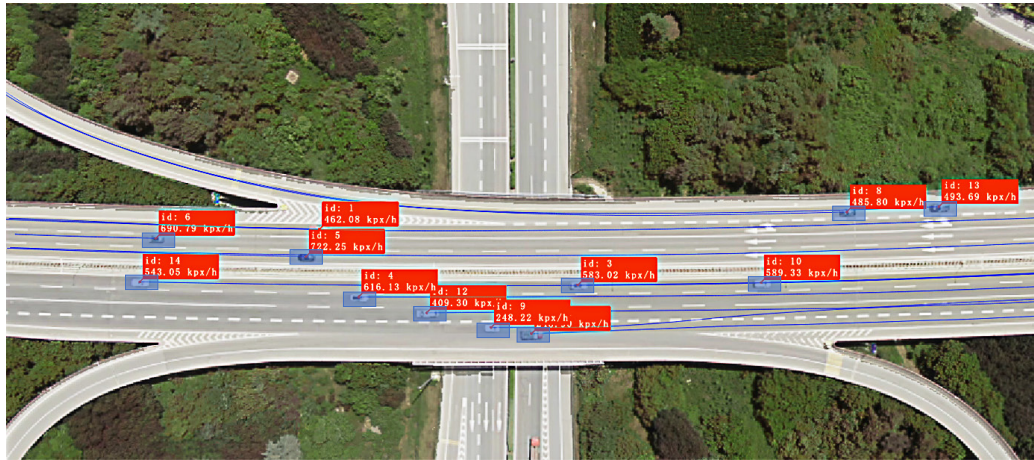


Figure 5 – Vehicle trajectory extraction interface in a weaving section

Table 3 – Sample trajectory data

Track ID	Type	Average speed [kpx/h]	X [px]	Y [px]	Speed [kpx/h]	Acceleration [px/s ²]	Time [s]
9	Car	426.515	1612.51	703.53	442.809	2.269	1.268
9	Car	426.515	1608.43	703.25	442.553	2.233	1.301
9	Car	426.515	1604.36	702.97	442.321	1.902	1.335
9	Car	426.515	1600.29	702.69	442.135	1.481	1.368
9	Car	426.515	1596.22	702.40	441.972	1.529	1.401

Firstly, it is necessary to extract trajectory data of lane-changing vehicles from the entire dataset to facilitate the analysis of lane-changing manoeuvres in subsequent sections of the paper. Secondly, it is essential to determine the lane location that lane-changing vehicles were in before and after lane-changing. This is achieved by comparing the X-values of trajectory lines with those of lane lines, with the intersection point between the trajectory line and lane line representing the lane-changing location, at which point their X-values are extremely close. Due to an inherent distortion of lane lines in the aerial videos, it was essential to fit the lane lines of the weaving sections. The curve estimation function of Statistical Product and Service Solutions (SPSS) was used to determine the model that best fits the data. According to fitting results, cubic polynomials had slightly smaller residuals than quadratic polynomials, with an R² of 0.997 for both, thus the former was chosen. The ANOVA SIG value is 0.00001, demonstrating its significance. Finally, 1,186 lane-changing trajectories were extracted, along with data pertaining to the lane-changing vehicle’s IDs, locations, frequencies of lane-changing and other information.

4.3 Analysis of variable importance results

The variables in the speed-influencing variable database proposed in this paper were ranked according to their importance, as shown in Figure 6. The significance of all variables is positive. This indicates that they all contribute positively to traffic speed. The results partially support the rationale for designing three new types of variables for weaving sections. Among all variables, vehicle spatial distribution variables made the biggest contribution. In comparison, the lane-changing manoeuvre variables contributed the least. The primary explanation for the limited contribution of lane-changing manoeuvres may be the small sample size, as lane-changing samples represent only about 3% of the total sample. In the experiments, the lane-changing manoeuvre variables were assigned values only for lane-changing samples and not for non-lane-changing samples, which explains their relatively lower overall contribution. Despite the small sample size, their positive impact

on the results is significant enough to warrant consideration in subsequent models. Another reason for considering the headway variable is that it is significantly correlated with the level of safety in the weaving section. The greater the headway between the target vehicle and the vehicle on the target lane before and after the vehicle, the greater the safety of the target vehicle after the lane change. Similarly, if the target vehicle is in the current lane, the greater the headway between the front and rear of the vehicle, the greater the degree of freedom of the driver to choose to change lanes and not easily change lanes in the process of colliding with surrounding vehicles. The contribution of the distance from the on-ramp variable is greater than the contribution of the distance from the off-ramp variable among the type of vehicles and on/off-ramp location relativity variables. This shows that, compared with the off-ramp, the on-ramp has a more substantial impact on driver travel speed. Traffic volume, large vehicle ratio, short length of weaving sections and number of lanes are all significant variables among the empirical variables affecting speed. The geometry of a weaving section is determined collectively by its short length and number of lanes, implying that the real operational speed condition of weaving sections can be improved from a design perspective. In conclusion, as every variable in the speed-influencing variable dataset positively affects speed, they are all included in the subsequent traffic speed prediction algorithm.

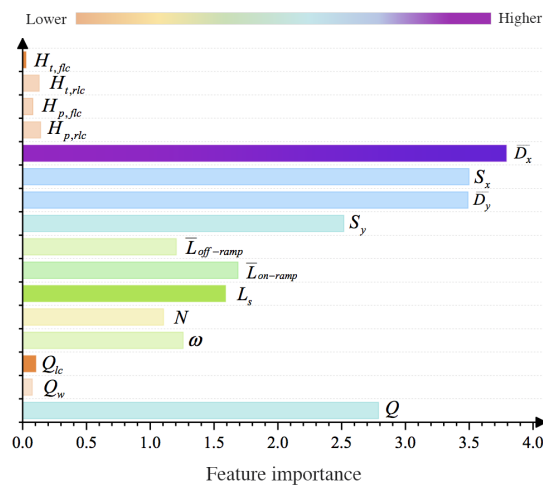


Figure 6 – Variable importance ranking

4.4 Comparison analysis of three machine learning algorithms

This paper initially employed three algorithms for selecting the weaving section speed prediction algorithm: RF, SVR and BPNN-GA. The best-performing algorithm was then selected for the speed prediction method. To increase processing speed, 12.5% of the total dataset samples were randomly chosen for the study. This study included a sample size of 10,892 items, with 321 items representing lane-changing samples. The ratio of a training set to a testing set was set at 4:1. R^2 , RMSE and MAE were chosen as the evaluation metrics for the regression algorithms to assess their effectiveness. The results are shown in Table 4.

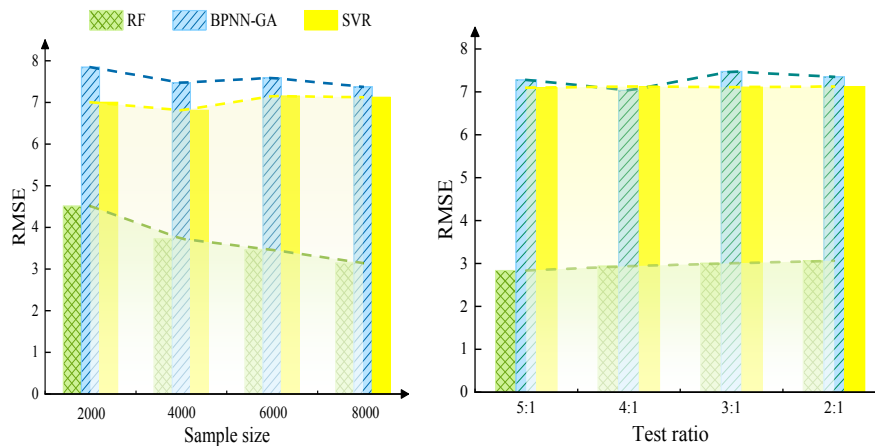
Table 4 – Regression evaluation indicators of each algorithm

Algorithm	Training set			Testing set		
	R^2	RMSE	MAE	R^2	RMSE	MAE
RF	0.959	2.939	1.911	0.915	4.443	2.943
BPNN-GA	0.762	7.034	5.413	0.751	7.397	5.070
SVR	0.761	7.124	5.068	0.746	7.270	5.266

R^2 reflects the extent to which the regression algorithm explains the actual values; RMSE indicates the magnitude of the deviation between predicted values and actual values; MAE represents the average of the absolute differences between predicted values and actual values, reflecting the actual error between them. The results of the training set reveal that the RF achieves the best regression performance, with a prediction error of

approximately two kilometres per hour compared to real speed. In contrast, the other two algorithms perform similarly, with prediction errors of approximately five kilometres per hour. The results of the test set exhibit a similar pattern to those of the training set, but the RF algorithm experiences a reduction in prediction accuracy of about one kilometre per hour.

To assess the adaptability of the proposed method across varying data input sizes, the method was tested to determine its capacity to perform well even with limited data inputs. This aligns with a recent research focus on achieving “high-accuracy prediction results with small sample inputs.” Thus, in this study, further experiments were conducted to investigate whether the data sample size and the ratio of training to testing sets had a significant impact on the results of the aforementioned method. From 10,892 raw data, samples with sizes of 2,000, 4,000, 6,000 and 8,000 were randomly selected for the experiment. Figure 7a displays the evaluation results using RMSE. The findings indicate that algorithm performances generally improve with increasing data volume, particularly when the overall data sample size is around 10,000. The performance remains stable and the algorithm’s effectiveness does not undergo abrupt change with varying samples. These experiments demonstrate that the data sample size does not significantly affect the algorithm’s performance. Notably, the RF algorithm consistently outperforms the other two algorithms.



a) RMSEs under different sample sizes
 b) RMSEs under different test ratios
 Figure 7 – RMSEs of the three algorithms under different sample sizes and test ratios

Secondly, the study aimed to verify whether the three algorithms consistently delivered stable performance across various test ratios. The experiments were conducted with training set to testing set ratios of 5:1, 4:1, 3:1 and 2:1, respectively. RMSE was used to assess the algorithm’s performance and the results are displayed in Figure 7b. The assessment metrics for each algorithm remain consistent across these four different test ratios, indicating that the choice of test ratios has minimal impact on the results.

Through two further experiments in this study, focusing on data sample size and test ratios, it was found that these two parameters barely interfere with the model’s prediction capacity, indirectly demonstrating the stability of the speed prediction method developed in this paper. The preferred prediction method, based on the RF algorithm, exhibits a speed prediction error of approximately three kilometres per hour. To the best of our knowledge, the accurate speed prediction error in the current research is about eight kilometres per hour [13]. These findings indicate that this study contributes to further enhancing the accuracy of spatial average speed prediction in weaving areas.

4.5 Comparison analysis of previous research methods

The speed prediction method proposed in this study was compared using HCM6 and LSTM models. We aggregated the extracted trajectory data into two categories, 5 minutes and 10 minutes, and compared the performance of three methods for estimating spatially averaged speeds, as shown in Table 5. The LSTM neural network used for comparison consists of an input layer, an LSTM layer, an activation layer, a fully connected layer and an output layer. The Adam gradient descent method was used, Mini Batch Size was set to 30, Max Epoch was set to 1,000, Initial Learn Rate was set to 1e-2 and Learn Rate Drop Factor was set

to 0.5. The HCM methodology can be specified in the weaving section of HCM6. The DTF-RF algorithm is the method proposed in this paper for weaving sections, DTF means Designing Three kinds of Features, and it can be seen that it demonstrates a higher estimation accuracy than LSTM and HCM6.

Table 5 – Regression evaluation indicators of each algorithm

Method	5 minutes		10 minutes	
	RMSE	MAE	RMSE	MAE
DTF-RF	2.94	1.91	3.23	1.78
LSTM	5.22	2.54	5.56	2.63
HCM6	35.42	16.73	40.23	19.62

5. CONCLUSION AND FUTURE WORK

This study contributes to speed prediction models for weaving sections by utilising traffic flow characteristics of weaving sections. The method adopts exploratory causation logic, distinguishing itself from the commonly used time series prediction logic in previous research. Specifically, we conducted an in-depth analysis of the speed distribution pattern in weaving sections and explored the unique speed-influencing factors. Furthermore, we integrated these factors with machine learning algorithms to construct this method. Lastly, the method was validated using aerial video data collected in the field, with a prediction error of approximately three kilometres per hour. The proposed method can provide guidance for the development of traffic control strategies related to safer speeds of vehicles in weaving sections. Additionally, this method can be incorporated into the development of HCM6. Several useful findings can be generated in this study.

The lane-changing distribution within weaving sections exhibits a pattern that is nearly inverse of spatial averaged speed distribution. Specifically, the number of lane changes in the first half of the weaving section is greater than in the second half, with the peak occurring near the first third of the on-ramp. Simultaneously, both on-ramps and off-ramps witnessed a sudden drop in speed despite variations in the geographical locations of different weaving sections, with the former pronouncing a notably significant deceleration.

This study finds that lane-changing manoeuvres, the relative positioning of vehicles concerning on/off-ramps and the spatial distribution of vehicles within the weaving sections significantly influence the spatial average speed. It is worth noting that short length, number of lanes, large vehicle ratios and total traffic volume all impact the spatial average speed of weaving sections, but the degree of impact decreases progressively.

This paper is limited to A-type weaving sections only and does not discuss other configurations. Due to the limited duration of aerial video collection, only weaving sections can be studied within a short period. It is not possible to consider the effects of different periods or traffic volumes.

Future work can be conducted by collecting all-weather video data to investigate the traffic flow operation of weaving sections during different periods and under varying traffic volumes. In addition, reconstructing spatiotemporal patterns of traffic congestion, predicting traffic accidents and implementing active traffic management strategies in weaving sections could be investigated. By designing well-considered road markings to guide vehicle trajectories and suggesting appropriate speeds, we can assist vehicles in making safer lane changes, ultimately enhancing the safety of vehicles passing through weaving sections.

ACKNOWLEDGMENTS

The research is supported by the Natural Science Basic Research Program of Shaanxi (No.2023-JC-YB-391) and the Key Research and Development Program of Shanxi (No.202102020101014).

REFERENCES

- [1] Prasetijo J, et al. Analysis of maximum weaving length and lane-changing rate for two-sided weaving section for federal road FT050 based on HCM 2010. *International Journal of Integrated Engineering*. 2018;10(2):56-60. DOI: 10.30880/ijie.2018.10.02.011.

- [2] Zhao CW, et al. Choice of lane-changing point in an urban intertunnel weaving section based on random forest and support vector machine. *Promet-Traffic & Transportation*. 2023;35(2):161-174. DOI: 10.7307/ptt.v35i2.60.
- [3] Liao YG, et al. Risk distribution characteristics and optimization of short weaving area for complex municipal interchanges. *Journal of Advanced Transportation*. 2021;2021:1-10. DOI: 10.1155/2021/5573335.
- [4] Ma CX, et al. A novel STFSA-CNN-GRU hybrid model for short-term traffic speed prediction. *IEEE Transactions on Intelligent Transportation Systems*. 2023;24(4):3728-3737. DOI: 10.1109/tits.2021.3117835.
- [5] Vlahogianni EI, Karlaftis MG, Golias JC. Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C-Emerging Technologies*. 2014;43:3-19. DOI: 10.1016/j.trc.2014.01.005.
- [6] Lighthill MJ, Whitham GB. On kinematic waves. II. A theory of traffic flow on long crowded roads. *Proceedings of The Royal Society A Mathematical Physical and Engineering Sciences*. 1955;229(1178):317-345. DOI: 10.1098/rspa.1955.0089.
- [7] Jiang H, et al. Short-term speed prediction using remote microwave sensor data: Machine learning versus statistical model. *Mathematical Problems in Engineering*. 2016;2016:1-13. DOI: 10.1155/2016/9236156.
- [8] Yang XX, et al. Evaluation of short-term freeway speed prediction based on periodic analysis using statistical models and machine learning models. *Journal of Advanced Transportation*. 2020;2020:1-16. DOI: 10.1155/2020/9628957.
- [9] Ye Q, Szeto WY, Wong SC. Short-term traffic speed forecasting based on data recorded at irregular intervals. *IEEE Transactions on Intelligent Transportation Systems*. 2012;13(4):1727-1737. DOI: 10.1109/tits.2012.2203122.
- [10] Williams BM, et al. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results. *Journal of Transportation Engineering*. 2003;129(6):664-672. DOI: 10.1061/(ASCE)0733-947X(2003)129:6(664).
- [11] Qi Y, Ishak S. A hidden Markov model for short term prediction of traffic conditions on freeways. *Transportation Research Part C-Emerging Technologies*. 2014;43:95-111. DOI: 10.1016/j.trc.2014.02.007.
- [12] Zhan AY, et al. A traffic flow forecasting method based on the GA-SVR. *Journal of High Speed Networks*. 2022;28(2):97-106. DOI: 10.3233/jhs-220682.
- [13] Meng XW, et al. D-LSTM: Short-term road traffic speed prediction model based on GPS positioning data. *IEEE Transactions on Intelligent Transportation Systems*. 2022;23(3):2021-2030. DOI: 10.1109/tits.2020.3030546.
- [14] Guo SN, et al. Deep spatial-temporal 3D convolutional neural networks for traffic data forecasting. *IEEE Transactions on Intelligent Transportation Systems*. 2019;20(10):3913-3926. DOI: 10.1109/tits.2019.2906365.
- [15] Yuan Y, et al. Macroscopic traffic flow modeling with physics regularized Gaussian process: A new insight into machine learning applications in transportation. *Transportation Research Part B-Methodological*. 2021;146:88-110. DOI: 10.1016/j.trb.2021.02.007.
- [16] Tian Y, Wei C, Xu D. Traffic flow prediction based on stack autoencoder and long short-term memory network. *2020 IEEE 3rd International Conference on Automation, Electronics and Electrical Engineering (AUTEEE), 20-22 November 2020, Shenyang, China*. 2021. p. 385-388. DOI: 10.1109/AUTEEE50969.2020.9315723.
- [17] Li LC, et al. Traffic speed prediction for intelligent transportation system based on a deep feature fusion model. *Journal of Intelligent Transportation Systems*. 2019;23(6):605-616. DOI: 10.1080/15472450.2019.1583965.
- [18] Song ZG, et al. Short-term traffic speed prediction under different data collection time intervals using a SARIMA-SDGM hybrid prediction model. *Plos One*. 2019;14(6):1-19. DOI: 10.1371/journal.pone.0218626.
- [19] Cao MM, Li V, Chan V. A CNN-LSTM model for traffic speed prediction. *IEEE 91st Vehicular Technology Conference (VTC2020-Spring), 25-28 May 2020, Antwerp, Belgium*. 2020. p. 1-5. DOI: 10.1109/VTC2020Spring48590.2020.9129440.
- [20] Niu K, et al. A novel spatio-temporal model for city-scale traffic speed prediction. *IEEE ACCESS*. 2019;7:30050-30057. DOI: 10.1109/access.2019.2902185.
- [21] Zhao YC, et al. Dynamic analysis of Kalman filter for traffic flow forecasting in sensor nets. *Information Technology Journal*. 2012;11(10):1508-1512. DOI: 10.3923/itj.2012.1508.1512.
- [22] Guo JH, Huang W, Williams BM. Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification. *Transportation Research Part C-Emerging Technologies*. 2014;43:50-64. DOI: 10.1016/j.trc.2014.02.006.
- [23] Cui Z, et al. Deep bidirectional and unidirectional LSTM recurrent neural network for network-wide traffic speed prediction. 2018:1-12. DOI: 10.48550/arXiv.1801.02143.

- [24] Ahmed I, et al. Lane change rates at freeway weaving sites: Trends in HCM6 and from NGSIM trajectories. *Transportation Research Record*. 2019;2673(5):627-636. DOI: 10.1177/0361198119841281.
- [25] Xu DZ, et al. Modeling framework for capacity analysis of freeway segments: Application to ramp weaves. *Transportation Research Record*. 2020;2674(1):148-159. DOI: 10.1177/0361198119900157.
- [26] Sun J, Jiang JR, Zheng JX. Improved speed prediction models on weaving segments of urban expressway. *China Journal of Highway and Transport*. 2017;30(01):83-91. DOI: 10.19721/j.cnki.1001-7372.2017.01.011.
- [27] Yu R, et al. Deep learning: A generic approach for extreme condition traffic forecasting. *Proceedings of the 2017 SIAM International Conference on Data Mining*. 2017. p. 777-785. DOI: 10.1137/1.9781611974973.87.
- [28] Ma XL, et al. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transportation Research Part C-Emerging Technologies*. 2015;54:187-197. DOI: 10.1016/j.trc.2015.03.014.
- [29] Dai XY, et al. Deeptrend 2.0: A light-weighted multi-scale traffic prediction model using detrending. *Transportation Research Part C-Emerging Technologies*. 2019;103:142-157. DOI: 10.1016/j.trc.2019.03.022.
- [30] Song C, et al. Traffic speed prediction under weekday using convolutional neural networks concepts. *28th IEEE Intelligent Vehicles Symposium (IV), 11-14 June 2017, Los Angeles, CA*. 2017. p. 1293-1298. DOI: 10.1109/IVS.2017.7995890.
- [31] Ma XL, et al. Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction. *Sensors*. 2017;17(4):1-16. DOI: 10.3390/s17040818.
- [32] Wang JY, et al. Traffic speed prediction and congestion source exploration: A deep learning method. *16th IEEE International Conference on Data Mining (ICDM), 12-15 December 2016, Barcelona, Spain*. 2017. p. 499-508. DOI: 10.1109/icdm.2016.51.
- [33] Ke RM, et al. Two-stream multi-channel convolutional neural network for multi-lane traffic speed prediction considering traffic volume impact. *Transportation Research Record*. 2020;2674(4):459-470. DOI: 10.1177/0361198120911052.
- [34] Fan Q, et al. Space-time hybrid model for short-time travel speed prediction. *Discrete Dynamics in Nature and Society*. 2018;2018:1-9. DOI: 10.1155/2018/7696592.
- [35] Tan MC, Feng YB, Xu JM. Traffic flow prediction based on hybrid ARIMA and ANN model. *China Journal of Highway and Transport*. 2007;(4):118-121. DOI: 10.19721/j.cnki.1001-7372.2007.04.023.
- [36] Wang HZ, et al. Empirical mode decomposition-autoregressive integrated moving average hybrid short-term traffic speed prediction model. *Transportation Research Record*. 2014;(2460):66-76. DOI: 10.3141/2460-08.
- [37] Wang HZ, et al. A novel work zone short-term vehicle-type specific traffic speed prediction model through the hybrid EMD-ARIMA framework. *Transportmetrica B-Transport Dynamics*. 2016;4(3):159-186. DOI: 10.1080/21680566.2015.1060582.

周郁茗, 张敏, 张驰, 谢子龙, 王博

一种新型交织区速度预测模型: 融合交通流特征

摘要

交织路段是车辆强制变道空间分布密集的重要区域, 极易引发交通事故。车速是衡量交通安全的基本指标, 因此开发精确的车速预测方法是提高交织路段安全性的关键一步。虽然现有方法在简单场景中表现出强大的预测能力, 但在复杂场景(如交织路段)中却显得力不从心。本研究基于上述挑战, 提出了一种针对交织路段的新型车速预测方法。首先, 开发了独特的变量以捕捉交织路段内区别于普通路段的交通流特征。其次, 纳入了已知影响车速的其他经验变量。随后, 进行了变量重要度评估, 以确定每个变量的贡献程度和方向。最后, 选择了具有显著正效应的变量作为三种机器学习算法的输入: 随机森林(RF)、遗传算法优化的反向传播神经网络(BPNN-GA)和支持向量回归(SVR)。该方法利用中国五个实际交织路段的航拍视频中进行了验证, 速度预测误差保持在约 3 公里/小时。此外, 研究还发现, 交织路段的车速分布与变道次数呈负相关。车辆在入口匝道和出口匝道都会减速, 但是前者的减速行为更显著。车速受交织区短长度、车道数和大型车比例的影响较

大。提出的方法可嵌入智能交通系统，以确保自动驾驶车辆在交织路段的安全行驶速度。进一步可研究交织路段交通拥堵时空模式重建、交通事故预测和主动交通管理策略效益。

关键词

交通安全，交织区，速度预测方法，车辆轨迹数据，变量重要度