Electronic Supplementary Materials

For https://doi.org/10.1631/jzus.A2500331

Lateral risk prediction and influencing factors analysis of container trucks based on trajectory reconstruction data

Zhihao ZHU, Hexuan LIU, Rongjun CHENG

Faculty of Maritime and Transportation, Ningbo University, Ningbo 315211, China

Section S1

$$ACC = \frac{TP + TN}{TP + FP + FN + TN} \tag{S1}$$

$$FNR = \frac{FN}{TP + FN} \tag{S2}$$

$$FPR = \frac{FP}{TN + FP} \tag{S3}$$

$$TPR = \frac{TP}{TP + FN} \tag{S4}$$

where TP is the correct prediction of the positive class, TN is the correct prediction of the negative class, FN is the incorrect prediction of the negative class as the positive class, and FP is the incorrect prediction of the positive class as the negative class.

$$f(x) = \phi_0 + \sum_{i=1}^{M} \phi_i \tag{S5}$$

where $f(x)$ is the model's predicted value for sample x, ϕ_0 is the model's output without any feature input (usually the average predicted value of all samples in the training set), ϕ_i is the marginal contribution of the feature i to the prediction result, that is the Shapley value of the feature, and M is the total number of features.

$$g(z^{'}) = \phi_0 + \sum_{i=1}^{M} \phi_i z_i^{'} \tag{S6}$$

where $Z_i' \in \{0,1\}^M$ represents how many features are included in the decision path of the sample.

For a sample, if the feature is not in its decision path, then $Z_i' = 0$. When all $Z_i' = 1$, the model degenerates to the actual predicted value $f(x)$. This model is a surrogate model constructed to calculate the Shapley value.

The Shapley value of each feature is defined as follows:

$$\phi_i = \sum_{S \subseteq M \setminus \{i\}} \frac{|S|!(|M|-|S|-1)!}{|M|!} \left[f_{S\cup\{i\}}(x_{s\cup\{i\}}) - f_s(x_s) \right] \tag{S7}$$

where S is the subset without feature i and $f_s(x_s)$ is the prediction for x_s by the model trained using only features from subset S.

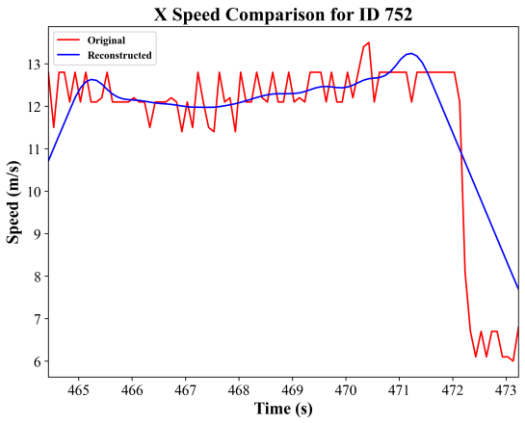
Section S2

Table S1 Introduction to raw data

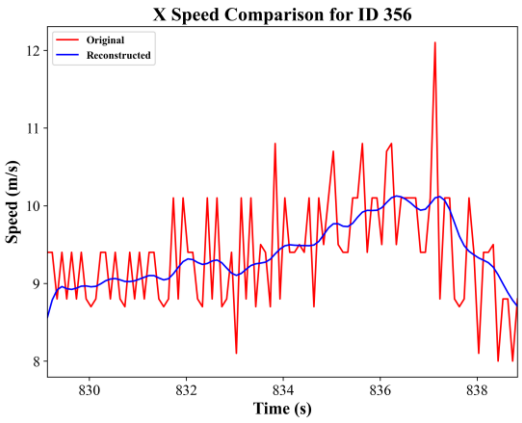
Raw data	description	Unit
Frame	Frame Time	–
ID	Vehicle ID	–
cls_Name	Vehicle Type: defines five different types of trucks, as well as a car type	–
(X_top_left,Y_top_left)	The coordinates of the top left corner of the vehicle detection box	–
(X_lower_right,Y_lower_right)	The coordinates of the lower right corner of the vehicle detection box	–
(X_center,Y_center)	The center coordinates of the vehicle	–
Length	Vehicle length	m
Width	Vehicle width	m
X_speed	Longitudinal speed of the vehicle	m/s
Y_speed	The lateral speed of the vehicle	m/s
X_acceleration	Longitudinal acceleration of the vehicle	m/s^2
Y_acceleration	The lateral acceleration of the vehicle	m/s^2

Section S3

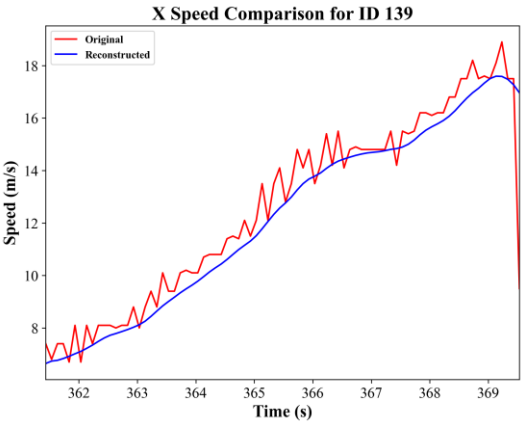
Note: The original data in Figs. S1–S5 refers to the trajectory data obtained through preliminary data preprocessing based on the raw data.



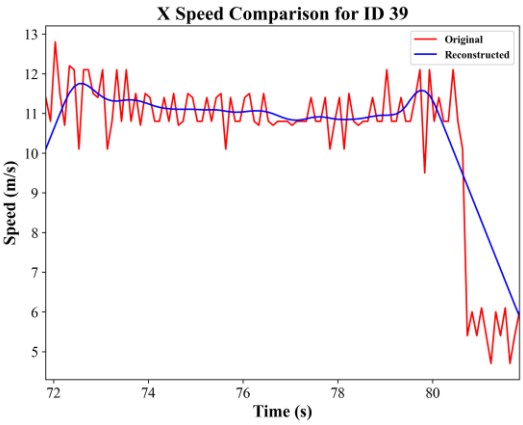
(a)



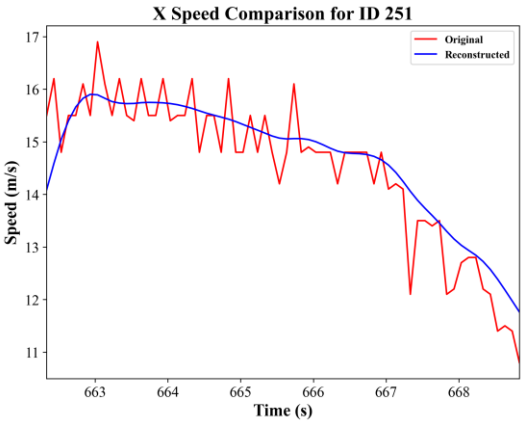
(b)



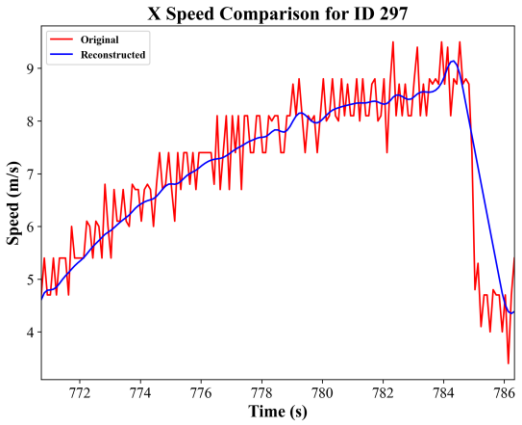
(c)



(d)



(e)



(f)

Fig. S1 Comparison of longitudinal velocity before and after reconstruction

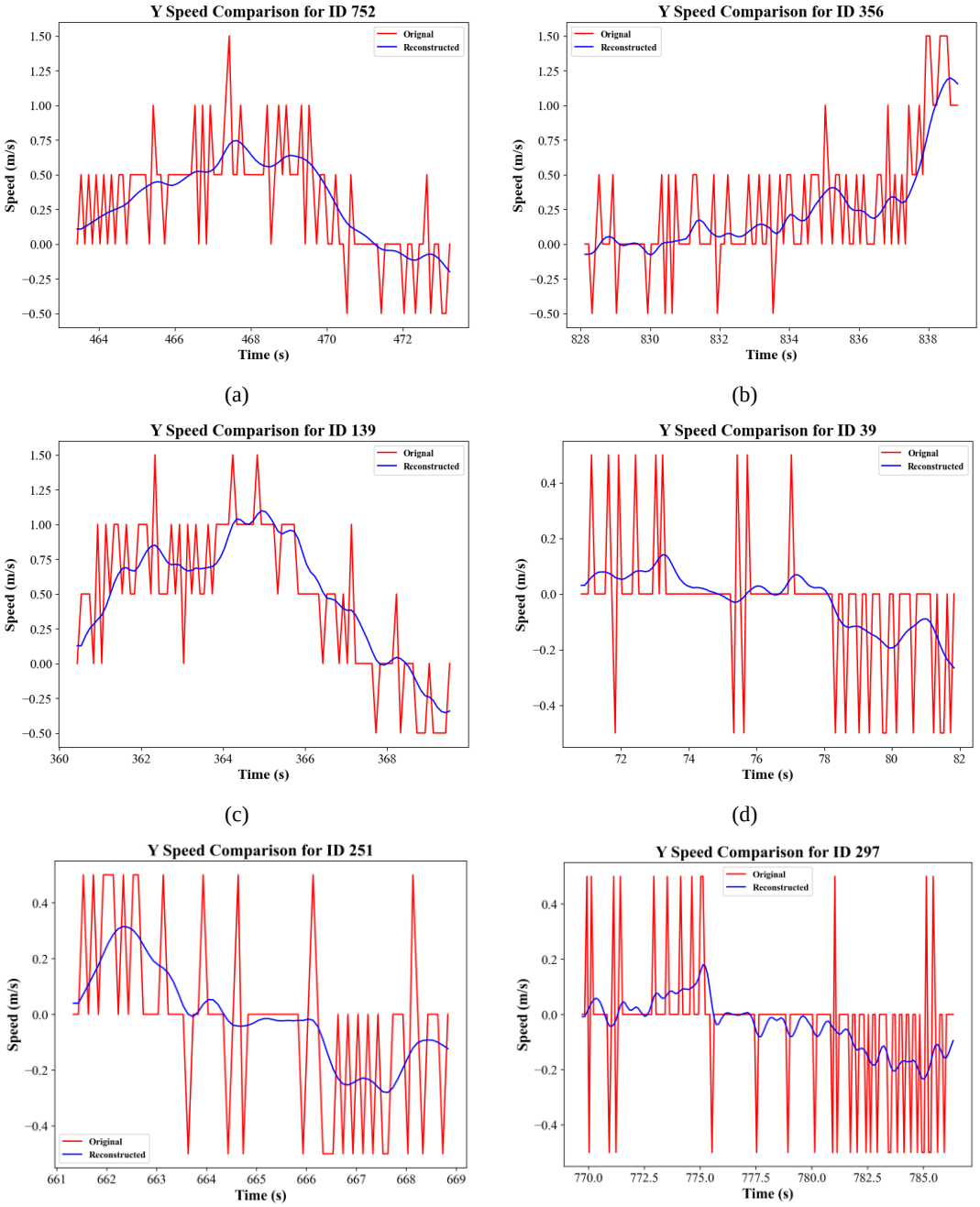
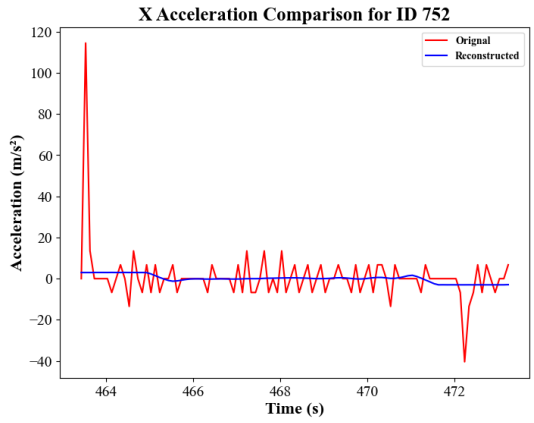
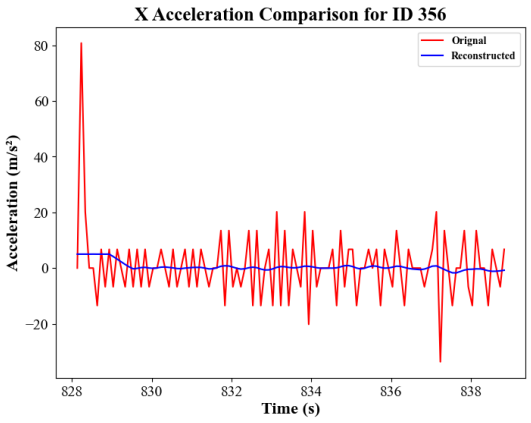


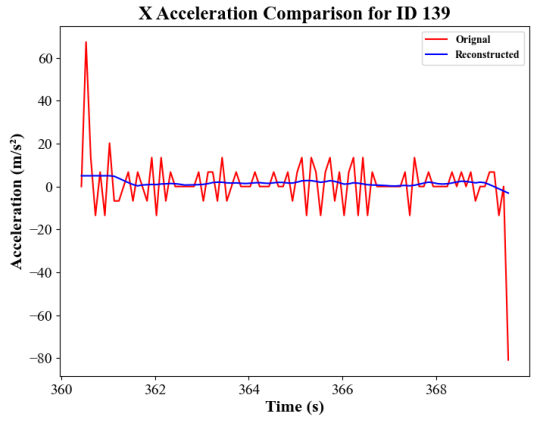
Fig. S2 Comparison of lateral velocity before and after reconstruction



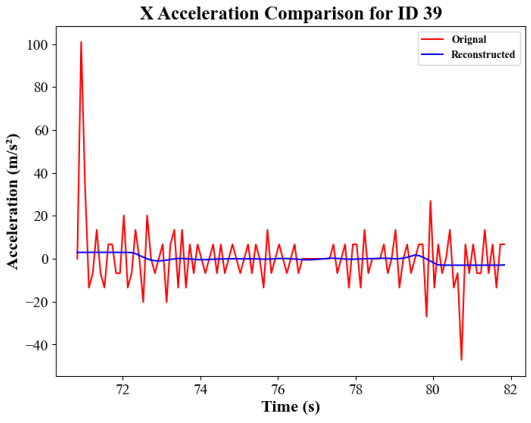
(a)



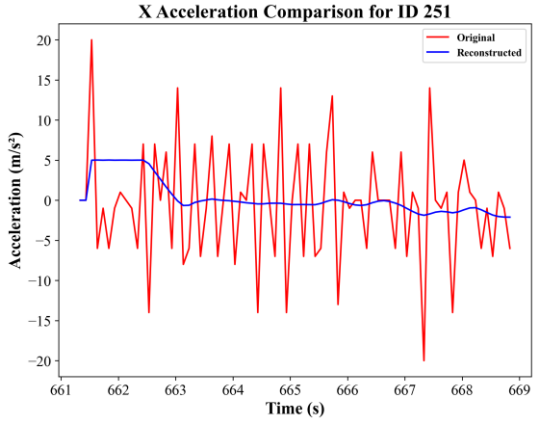
(b)



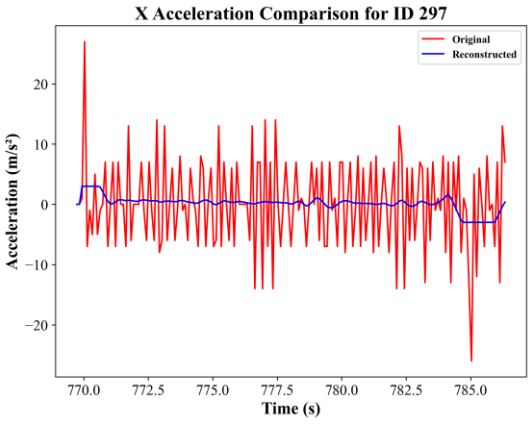
(c)



(d)

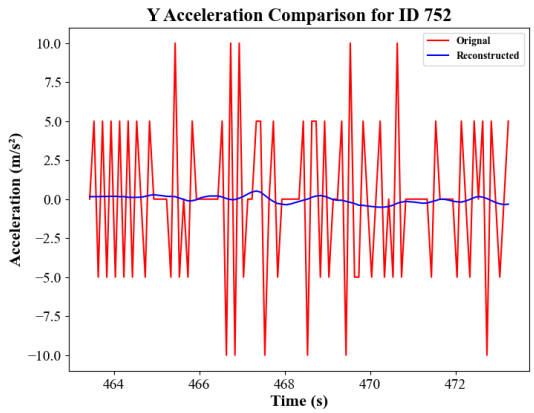


(e)

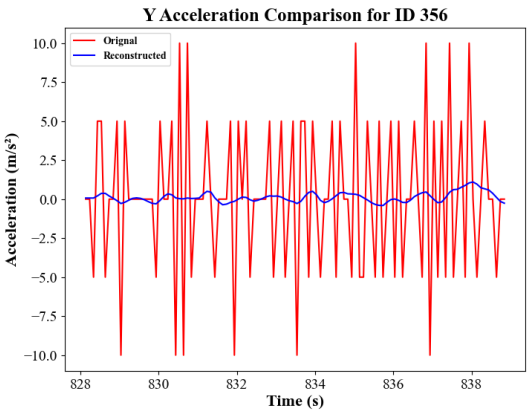


(f)

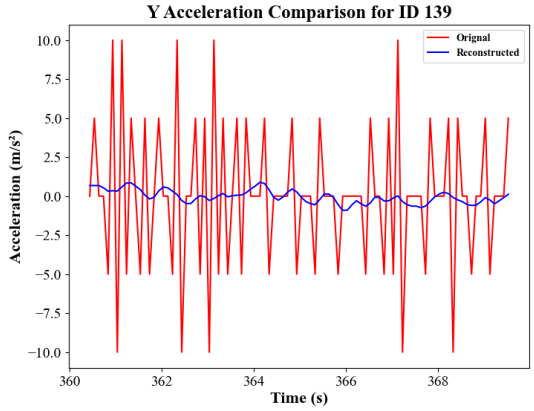
Fig. S3 Comparison of longitudinal acceleration before and after reconstruction



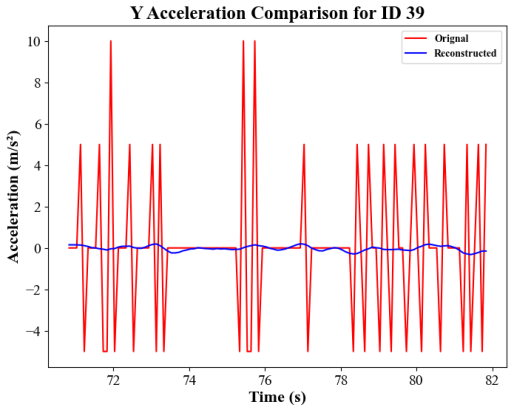
(a)



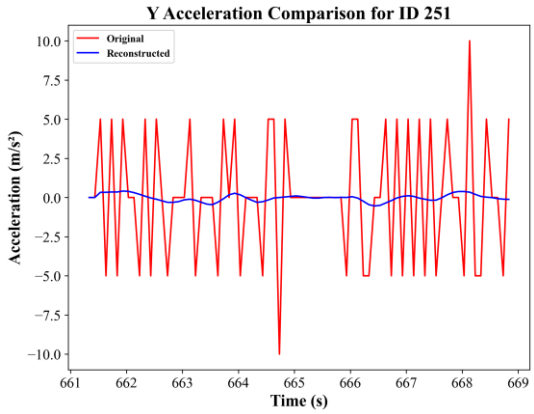
(b)



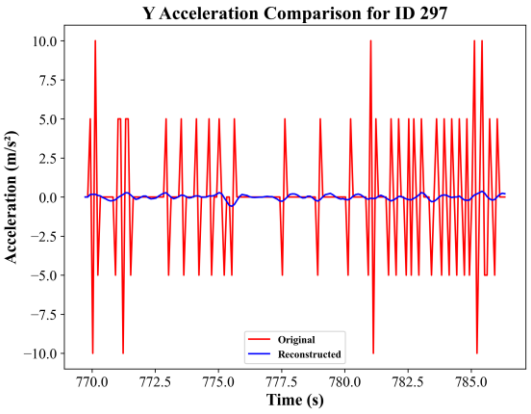
(c)



(d)

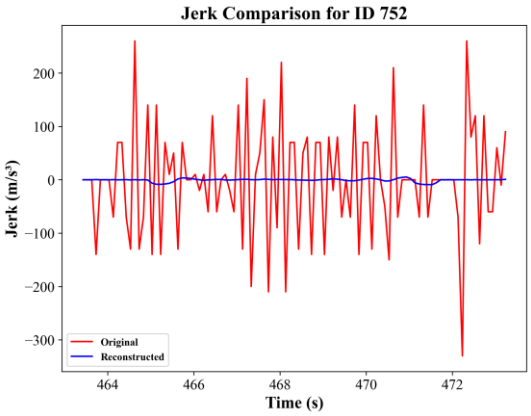


(e)

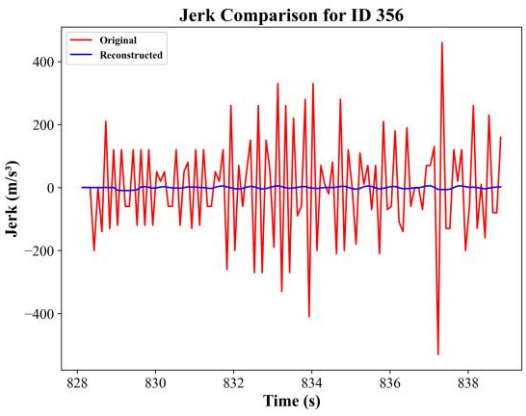


(f)

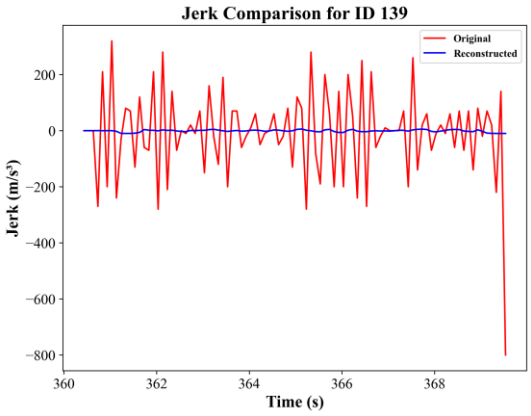
Fig. S4 Comparison of lateral acceleration before and after reconstruction



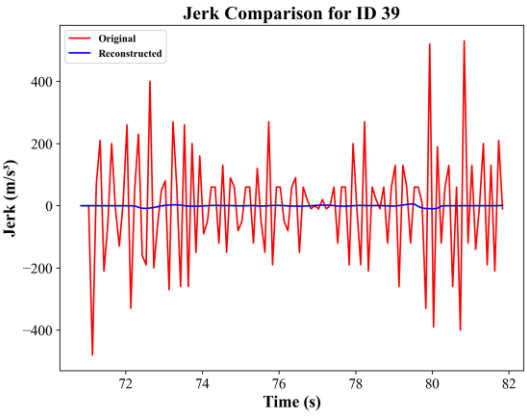
(a)



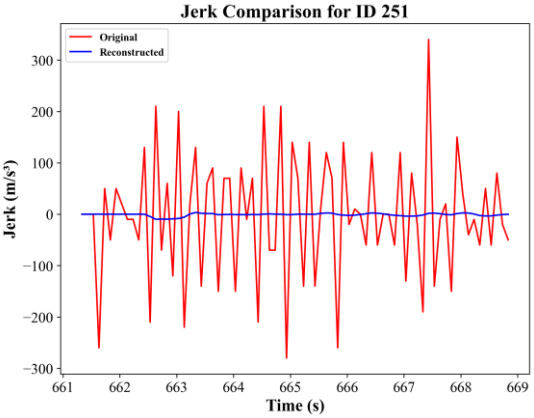
(b)



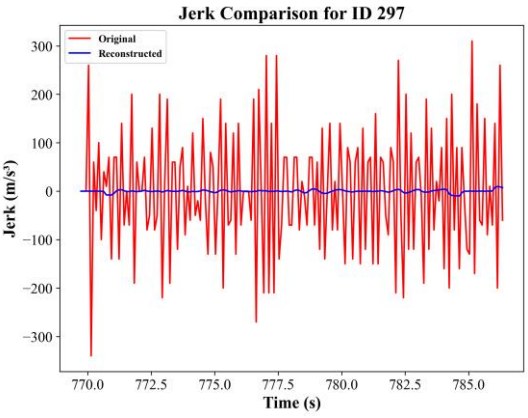
(c)



(d)



(e)



(f)

Fig. S5 Comparison of Jerk before and after reconstruction

Section S4

Table S2 Descriptive statistics of variable and conflict data (side-swipe)

	Variable	Unit	Decription	Mean	Min	Max	Std
A	X_speed_actual	m/s	Longitudinal speed	7.1367	0.82	18.57	3.2818
	Y_speed_actual	m/s	Lateral speed	0.1711	0.11	0.67	0.0968
	Heading_Angle	degree	Heading Angle	1.6966	0.3716	10.8347	1.2581
	X_acceleration_actual	m/s^2	Longitudinal acceleration	0.6367	-4.99	5.00	2.2064
	Y_acceleration_actual	m/s^2	Lateral acceleration	0.0548	-0.91	1.87	0.2998
B	avg_speed_x	m/s	Average longitudinal speed within 5s	6.9745	0.0847	15.15	3.1043
	avg_speed_y	m/s	Average lateral speed within 5s	0.1272	0.0022	0.9539	0.1203
	avg_accel_x	m/s^2	Average longitudinal acceleration within 5s	1.6125	-1.9624	5.00	1.8415
	avg_accel_y	m/s^2	Average lateral acceleration within 5 seconds	0.0976	-0.41	1.87	0.1761
C	std_speed_x	/	Standard deviation of longitudinal velocity within 5s	1.0515	0	3.1827	0.5547
	std_speed_y	/	Standard deviation of lateral velocity within 5s	0.0691	0	0.5227	0.0461
	std_accel_x	/	Standard deviation of longitudinal acceleration within 5s	1.0961	0	3.5826	0.9843
	std_accel_y	/	Standard deviation of lateral acceleration within 5s	0.1972	0	1.3402	0.1295
	std_heading_angle	/	Standard deviation of heading angle within 5s	0.6162	0	4.8924	0.4439
	TTC	s	Time to collsion	7.7654	0.88	34.09	4.7590

Section S5

Table S3 Descriptive statistics of variable and conflict data (rear-end)

	Variable	Unit	Decription	Mean	Min	Max	Std
A	X_speed_actual	m/s	Longitudinal speed	2.7497	0.5	15.55	2.3736
	Y_speed_actual	m/s	Lateral speed	0	0	0	0
	Heading_Angle	degree	Heading Angle	0	0	0	0
	X_acceleration_actual	m/s^2	Longitudinal acceleration	0.0343	-5.00	5.00	1.4096
	Y_acceleration_actual	m/s^2	Lateral acceleration	-0.0019	-1.03	2.26	0.162
B	avg_speed_x	m/s	Average longitudinal speed within 5s	3.3233	0.0188	14.45	2.5175
	avg_speed_y	m/s	Average lateral speed within 5s	0.0196	0	0.7557	0.0571
	avg_accel_x	m/s^2	Average longitudinal acceleration within 5s	0.2422	-2.9955	5.00	1.2290
	avg_accel_y	m/s^2	Average lateral acceleration within 5 seconds	-0.0014	-0.98	2.26	0.0774
C	std_speed_x	/	Standard deviation of longitudinal velocity within 5s	0.7743	0	4.0626	0.7743
	std_speed_y	/	Standard deviation of lateral velocity within 5s	0.0218	0	0.4169	0.0469
	std_accel_x	/	Standard deviation of longitudinal acceleration within 5s	0.6590	0	4.1685	0.6158
	std_accel_y	/	Standard deviation of lateral acceleration within 5s	0.1310	0	0.8916	0.1108
	std_heading_angle	/	Standard deviation of heading angle within 5s	0.2553	0	6.8540	0.5148
	TTC	s	Time to collsion	21.40	-1394	3771	155.9

Section S6

Table S4 Feature importance ranking

Feature	Ranking by importance	
	Side-swipe	Rear-end
X_speed_actual	12	3
Y_speed_actual	1	9
Heading_Angle	4	8
X_acceleration_actual	8	1
Y_acceleration_actual	14	7
avg_speed_x	**2**	**4**
avg_speed_y	13	6
avg_accel_x	**3**	**2**
avg_accel_y	10	5
std_speed_x	9	/
std_speed_y	11	/
std_accel_x	6	/
std_accel_y	5	/
std_heading_angle	7	/

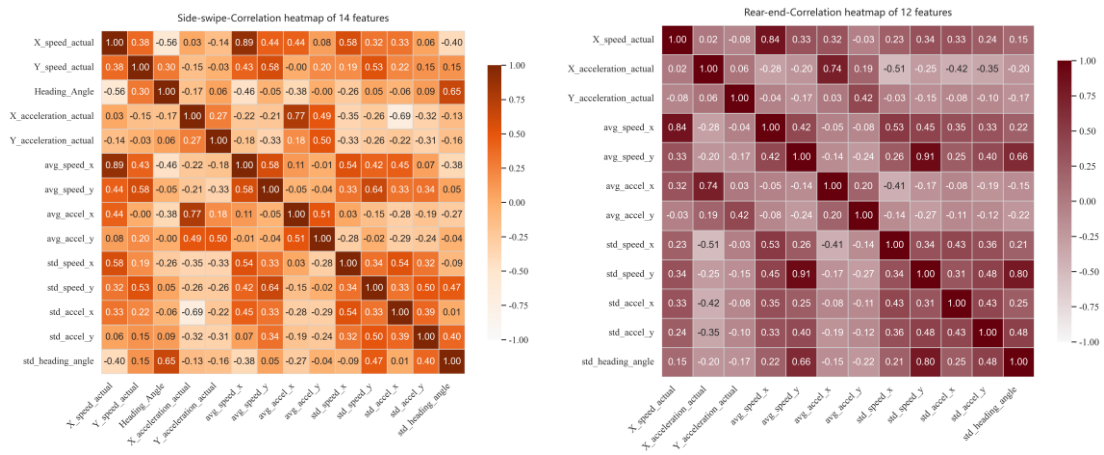
Section S7

Table S5 Experimental parameter settings

Conflict Type	Model	Model parameters					
Side-swipe	GBDT	learning_rate	max_depth	n_estimators	subsample	min_samples_leaf	min_samples_split
		0.2	6	100	1.0	1	10
Rear-end	XGBoost	learning_rate	max_depth	n_estimators	subsample	colsample_bytree	
		0.01	3	500	0.8	0.8	

Section S8: Feature correlation analysis

In the previous section on feature selection, we omitted feature correlation analysis. The original intention was to use SHAP analysis to rank feature importance and eliminate unnecessary features to optimize model prediction performance. However, highly correlated features can significantly impact models like LR and SVM. Therefore, this section discusses whether LR and SVM models outperform other tree-based models after removing highly correlated features.



(a) (b)

Fig. S6 Heatmap of feature correlation between two conflict types: (a)side-swipe; (b)rear-end

Fig.S6 shows a heat map of feature correlations between side-swipe conflicts and rear-end conflicts. We removed highly correlated features and fed the remaining features into the model for training and testing using the same experimental environment as previously described. The remaining features are shown in Table S6.

Table S6 Retained features

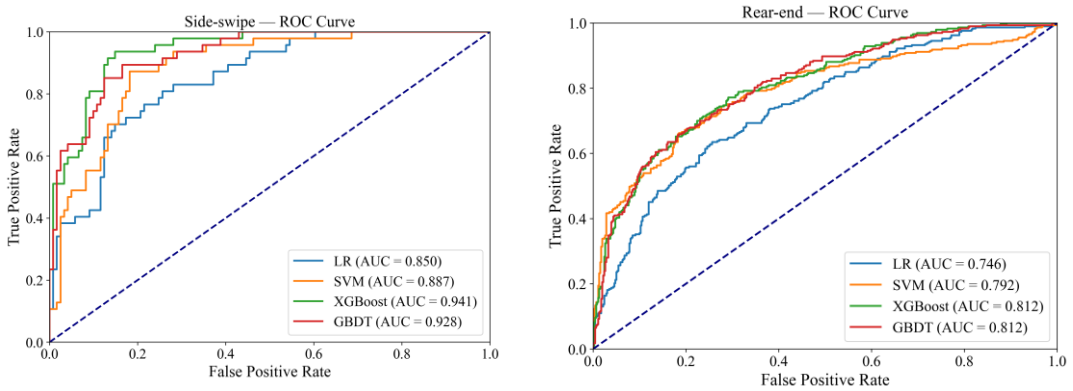
Feature	Retained features	
	Side-swipe	Rear-end
X_speed_actual	√	√
Y_speed_actual	√	
Heading_Angle	√	
X_acceleration_actual	√	√
Y_acceleration_actual	√	√
avg_speed_x		
avg_speed_y		√
avg_accel_x		
avg_accel_y		
std_speed_x	√	√
std_speed_y	√	
std_accel_x	√	
std_accel_y	√	√
std_heading_angle	√	√

The retained features were input into the machine learning model again, and the parameter grid tuning settings were consistent. The final prediction results of the two conflict types are shown in Table S7. The ROC curves of each model are shown in Fig. S7.

Table S7 Comparison of conflict prediction results (features retained)

	Model	ACC	FPR	FNR
Side-swipe	LR	0.774	0.091	0.574
	SVM	0.804	0.083	0.489
	XGBoost	**0.881**	**0.074**	**0.234**
	GBDT	0.851	0.099	0.277
Rear-end	LR	0.705	0.176	0.481
	SVM	0.754	**0.100**	0.474
	XGBoost	0.758	0.122	**0.430**
	GBDT	**0.761**	0.111	0.440

By comparing Table S7 with Table 2 and Table 3; Fig. S7 and Fig. 8, we can see that the performance of the model trained by inputting the retained features into the machine learning model is not as good as that of the model trained by inputting multiple features. The purpose of this paper is to develop a real-time conflict prediction model with good predictive performance, so it is reasonable to use all these variables for modeling. Finally, this paper conducts a feature ablation experiment based on the feature importance ranking to improve the model performance, which is also a way to weaken this effect.



(a) (b)

Fig. S7 ROC curves of different models (features retained): (a)side-swipe; (b)rear-end