J Intel Connect Veh

Ego vehicle trajectory prediction based on driver-vehicle coupling characteristics in diverse urban scenarios

Guoying Chen[1], Zheng Gao[2], Tianjun Sun[1,✉], Xinyu Wang[2,✉], Min Hua[3], Zhenhai Gao[1]

Cite this article: Chen G Y, Gao Z, Sun T J, et al. J Intel Connect Veh 2026, **9**(2): 9210082. https://doi.org/10.26599/JICV.2026.9210082

ABSTRACT: For human–machine co-driving vehicles, the dynamic complexity of urban scenarios and the uncertainty of driving behavior impose stringent demands on the accuracy and generalization capability of ego vehicle trajectory prediction. Current research predominantly relies on multivehicle interaction information or single-scenario settings, overlooking the inherent dynamic correlation between driver and vehicle. This results in prediction models struggle to adapt to complex urban environments. To address this, this study proposes a trajectory prediction framework based on driver-vehicle coupling feature encoding, enabling precise capture of vehicle trajectory evolution patterns under driver manipulation across diverse urban scenarios. The framework employs bidirectional long short-term memories (LSTMs) to perform temporal encoding on historical driver-vehicle coupling features and future road geometric features. Combined with an attention mechanism, it generates context vectors integrating temporal features and manipulation details, ultimately using LSTMs to recursively produce multistep prediction results. Validation using diverse urban scenarios data from a dynamic driving simulatordemonstrates that our prediction framework achieves precise trajectory prediction in typical scenarios such as lane changing, turning, and roundabout navigation, while exhibiting robust stability and generalization capabilities.

KEYWORDS: trajectory prediction; human–machine co-driving vehicle; driver-vehicle coupling system; diverse urban scenarios

1 Introduction

To address traffic accidents and congestion, intelligent driving technology has garnered significant attention from researchers (Camarcat et al., 2025; Hussain et al., 2025). However, constrained by limitations in environmental perception accuracy and the ability to respond to sudden risks, current intelligent systems cannot yet provide zero-error driving safety assurance (Liu and Feng, 2024). On the path toward achieving safe and efficient transportation, human–machine cooperative driving remains the prevailing solution (Huang et al., 2024).

For human–machine co-driving vehicles, the primary driving tasks are performed by the driver (Huang et al., 2025). The intelligent system continuously monitors driving conditions and intervenes to correct operations when potential risks arise, enabling a seamless transfer of driving authority (Fang et al., 2023). The core prerequisite for this process is the precise prediction of vehicle trajectories—only by clearly defining the vehicle's future path can the intelligent system fully comprehend the driver's operational intent and driving risks, thereby facilitating a smoother and safer transfer of control between the driver and the intelligent system (Ansari et al., 2022; Huang et al., 2025).

Urban environments represent the core application scenario for human–machine cooperative driving. Characterized by complex road structures, unpredictable behaviors of traffic participants, and highly variable driving conditions, the vehicle trajectory in such settings is significantly influenced by the coupling effects of driver operations, vehicle dynamics, and road constraints (Hu et al., 2026; Paparusso et al., 2023; Sun et al., 2021; Zhang et al., 2024b). Consequently, trajectory prediction in urban settings proves far more challenging than in structured environments such as highways. Therefore, developing high-precision vehicle trajectory prediction methods tailored to urban scenarios holds significant theoretical and practical value for enhancing the high-level coordination of human–machine intentions (Tan et al., 2022).

Research on vehicle trajectory prediction has undergone long-term development, evolving from early physics-based methods to recent data-driven approaches. Relevant reviews have systematically outlined the evolution of various technical pathways (Huang et al., 2022; Mozaffari et al., 2022; Xing et al., 2019). Physics-based methods describe trajectory evolution by establishing vehicle kinematic or dynamic models, with their core principle being the use of physical laws to characterize vehicle motion patterns (Reisinger et al., 2021). Although straightforward to implement, extensive research indicates that such methods yield only satisfactory short-term prediction performance (less than 1 s) and struggle to adapt to long-term trajectory prediction (Zhang et al., 2024b). Consequently, recent research has primarily focused on data-driven approaches, including machine learning-based methods and deep learning-based methods. These techniques analyze motion, road, and interaction information embedded within input features to achieve single-modal, multimodal, or driving intent prediction for vehicles (Gao et al., 2025;

[1] State Key Laboratory of Automotive Chassis Integration and Bionics, Jilin University, Changchun 130022, China. [2] College of Automotive Engineering, Jilin University, Changchun 130022, China. [3] School of Engineering, University of Birmingham, Birmingham B15 2TT, UK.

✉ Corresponding authors. E-mail: T. Sun, sun_tj@jlu.edu.cn; X. Wang, wangxinyu21@mails.jlu.edu.cn

Received: January 6, 2026; Revised: March 2, 2026; Accepted: March 21, 2026

Zhang et al., 2024b).

Machine learning-based methods utilize different motion models to represent distinct trajectory segments, achieving trajectory prediction through behavior matching. Common approaches include the Gaussian process, dynamic Bayesian network, and hidden Markov model (Chen et al., 2024; Gao et al., 2023; Jiang et al., 2023; Liu et al., 2020). Although these methods perform well in long-term predictions, they exhibit lower accuracy in short-term predictions and suffer from computational inefficiency (Schuetz and Flohr, 2024; Bharilya and Kumar, 2024). Some researchers have proposed integrating physical models with machine learning methods to enhance the overall prediction performance across the temporal domain. However, such fusion approaches still suffer from computational inefficiency (Bikmukhametov and Jäschke, 2020; Chen et al., 2024; Muther et al., 2023).

Deep learning-based methods have emerged as a research hotspot in trajectory prediction due to their powerful nonlinear fitting and temporal feature extraction capabilities. Early studies employed models such as long short-term memory (LSTM) and gate recurrent units (GRUs) to capture temporal dependencies (Liao et al., 2024; Meng et al., 2023; Wang et al., 2024, 2025b). Subsequently, the field evolved toward the encoder-decoder paradigm, where Transformers further enhanced the ability to capture long-term sequential dependencies through self-attention mechanisms (Geng et al., 2023; Giuliari et al., 2021; Yang et al., 2022; Zhang et al., 2025). Recent studies have integrated deep learning with physical equations to establish physics-informed neural network models, achieving a deep fusion of data-driven approaches with physical constraints (Zhang et al., 2024a). Additionally, language models are gradually finding application (Liu et al., 2023).

However, current deep learning-based research primarily utilizes open-source datasets to develop predictive models for specific scenarios, such as high-speed lane changing and merging at intersections. The behaviors within these datasets tend to be standardized, making it difficult to reflect drivers' personalized operations and complex decision-making logic in urban environments, which limits the models' generalization capabilities (Fang et al., 2024; Hu et al., 2022; Wang et al., 2025a). Furthermore, in pursuit of high-precision prediction performance, these approaches often fall into the dilemma of overly complex network architectures and inefficient training data utilization, struggling to meet the real-time decision-making demands of diverse urban scenarios (Hu et al., 2022; Wang et al., 2025b).

More notably, existing studies excessively focus on road environments and multivehicle interaction information, neglecting the core leading role of the driver-vehicle coupling system in trajectory generation and failing to deeply explore critical information within human–machine interactions (Tan et al., 2022).

In human–machine co-driving vehicles, the driver serves as the core decision maker for vehicle motion, following a clear logic-based closed-loop trajectory decision process: The driver first integrates personal driving experience with comfort requirements, comprehensively evaluating environmental constraints such as road geometry and traffic signals (Tan et al., 2022). Subsequently, by manipulating components such as the steering wheel, accelerator pedal, and brake pedal, the driver outputs operational commands. These commands directly influence the vehicle, inducing dynamic changes in its kinematic and dynamic states, ultimately forming the actual driving trajectory. This complete closed-loop system, encompassing environmental perception, driver decision-making, vehicle execution, and state feedback, embodies the intrinsic logic governing trajectory evolution. Existing studies, however, lack in-depth exploration of the interactive information between driver operational behavior and vehicle dynamic states within this closed-loop system (Tan et al., 2022). Specifically, there is no systematic integration of the dynamic relationship between driver manipulations and vehicle motion under road geometric constraints. This results in models struggling to accurately capture the trajectory variation patterns driven by the driver, severely limiting their generalization capabilities and prediction accuracy in complex, diverse urban scenarios (Fang et al., 2024).

To address these challenges, this study achieves trajectory prediction by encoding the dynamic characteristics of the driver-vehicle coupling system, termed driver-vehicle encoded trajectory predicting (DriVETP), as illustrated in Fig. 1. This algorithm focuses on driver-vehicle coupling features and road geometric constraints to precisely capture the core closed-loop logic of trajectory generation, enhancing the model's understanding of driver intent. First, a coupled feature system encompassing vehicle kinematics, dynamics, and driver manipulation parameters is constructed. Combined with future road geometric features, this system comprehensively captures the core factors influencing trajectory generation. Second, a lightweight network architecture featuring bidirectional LSTM (BiLSTM) encoding, attention-weighted processing, and LSTM decoding is designed. BiLSTM extracts temporal dependencies in features, while the attention mechanism dynamically amplifies the contribution of key driving behaviors to trajectory prediction. This achieves both temporal

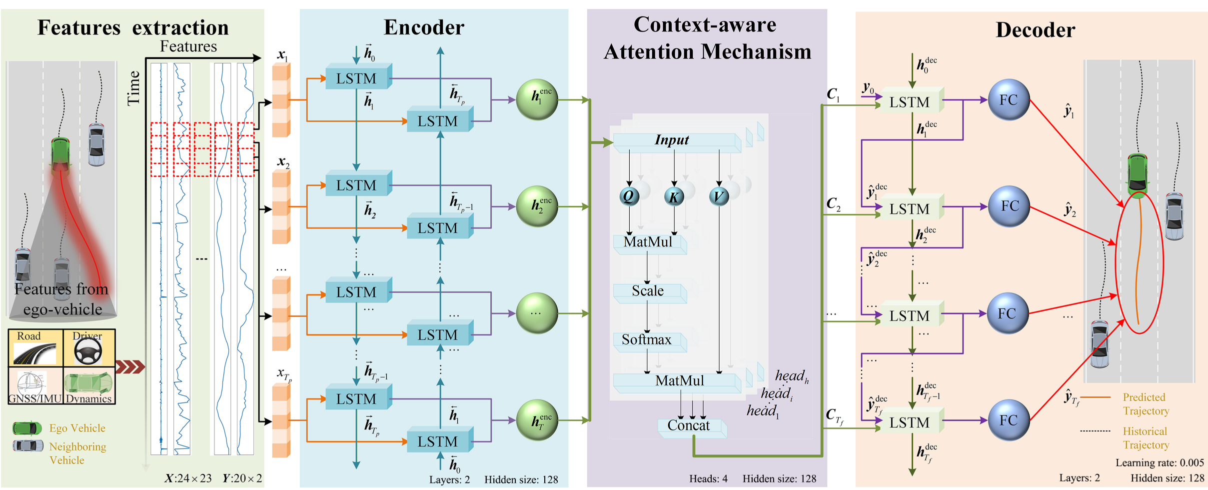

Fig. 1 Schematic of the proposed deep learning trajectory prediction framework based on driver-vehicle coupling feature encoding.

feature extraction and cross-modal feature fusion, ultimately generating multistep predicted trajectories through LSTM decoding. Experimental results demonstrate that despite its relatively simple network structure, DriVETP achieves high-precision deterministic short-to-medium-term prediction across diverse urban driving scenarios while maintaining strong generalization under varying manipulations. Our key innovations are summarized as follows:

- We propose the DriVETP framework, which systematically integrates driver-vehicle coupling characteristics to precisely capture the core closed-loop logic governing trajectory generation, thereby enhancing the model's ability to understand driver intent.
- We develop a lightweight prediction architecture that efficiently extracts bidirectional temporal features through BiLSTM while dynamically focusing on critical driving manipulations via attention mechanisms. This architecture achieves real-time, precise predictions across diverse urban scenarios.
- We conduct a comprehensive evaluation of DriVETP's predictive performance across diverse urban scenarios. Through comparative analysis with mainstream models and ablation studies, we systematically verify the proposed framework's prediction accuracy, real-time performance, and generalization capability.

The remainder of this paper is organized as follows. Section 2 provides a formal definition of the ego vehicle trajectory prediction problem. Section 3 details the network architecture of DriVETP. Section 4 presents the driving data collected via a dynamic driving simulator and validates DriVETP through comparisons with mainstream models and ablation experiments. Section 5 gives the conclusion.

2 Problem formulation

Trajectory prediction forms the foundation of decision-making, planning, and control in intelligent systems. Its essence lies in modeling the dynamic interactions between drivers, vehicles, and roads to achieve precise inference of a vehicle's future motion state. In human–machine co-driving scenarios, this task faces two core challenges. First, road geometric features dependent on real-time sensing are susceptible to sensor noise interference. Additionally, the limited scene coverage of high-definition maps and lane-level navigation maps leads to insufficient input features when lane line recognition fails, compromising model prediction stability (Messaoud et al., 2025). Second, existing research predominantly focuses on vehicle position and velocity features, overlooking the intrinsic coupling mechanism between driver manipulation and vehicle dynamic response (Tan et al., 2022). This causes predicted trajectories to overrely on kinematic constraints, resulting in overly conservative performance during urban sharp turning and consecutive lane changing scenarios.

To comprehensively capture the core factors influencing trajectory generation and address the issue of biased feature capture in existing research, this study constructs a 3-dimensional coupled feature system encompassing "driver-vehicle-road". All features from the historical T_p time windows are defined as the input feature matrix $\boldsymbol{X} = [\boldsymbol{x}_{t-T_p+1}; \cdots; \boldsymbol{x}_\tau; \cdots; \boldsymbol{x}_t] \in \mathbb{R}^{T_p \times n}$, where $n = 23$ represents the input feature dimension, and $\boldsymbol{x}_\tau = [\boldsymbol{s}^k, \boldsymbol{s}^c, \boldsymbol{s}^r]$ denotes the coupled feature vectors at moment τ. Specifically:

- Spatial kinematic features $\boldsymbol{s}^k \in \mathbb{R}^3$: Intuitively represent the vehicle's current spatial motion state, including the center-of-gravity (CG) position (x_o, y_o) and heading angle ϕ;
- Driver-vehicle coupling features $\boldsymbol{s}^c \in \mathbb{R}^{12}$: Composed of 8-dimensional vehicle dynamics features and 4-dimensional driver manipulation features, forming the core elements representing the manipulation-dynamics relationship. The dynamic features include CG velocities (v_x, v_y), vehicle slip angle β, yaw rate $\dot{\phi}$, and front/rear tire slip angles α_{FL}, α_{FR}, α_{RL}, α_{RR}, reflecting the vehicle's dynamic response to manipulation inputs. Manipulation fetures encompass the steering wheel angle δ_{SW}, steering wheel angular velocity $\dot{\delta}_{\mathrm{SW}}$, and longitudinal/lateral acceleration (a_x, a_y), directly mapping the driver's decision intent;
- Road geometric features $\boldsymbol{s}^r \in \mathbb{R}^8$: Constraints for the future 120 m road segment defining the physical boundaries for vehicle travel. By default, three lanes are formed by four lane lines, which are composed of the lane line coordinate sequences shown in Fig. 2. Missing lane lines will be filled with road boundary lines.

The coupling relationship between driver manipulation and vehicle dynamic response constitutes the core mechanism of trajectory evolution. Among these, the tire slip angle serves as the key parameter linking manipulation inputs to the vehicle dynamic response (Zhao et al., 2024), defined as in Eq. (1):

$$
\begin{aligned}
\alpha_{\mathrm{FL}} &= \arctan\frac{v_y + a\dot{\phi}}{v_x - \dot{\phi}w/2} - \delta \\
\alpha_{\mathrm{FR}} &= \arctan\frac{v_y + a\dot{\phi}}{v_x + \dot{\phi}w/2} - \delta \\
\alpha_{\mathrm{RL}} &= \arctan\frac{v_y - b\dot{\phi}}{v_x - \dot{\phi}w/2} \\
\alpha_{\mathrm{RR}} &= \arctan\frac{v_y - b\dot{\phi}}{v_x + \dot{\phi}w/2}
\end{aligned}
\tag{1}
$$

where a and b denote the distances from the vehicle's CG to the front and rear axles, respectively. w denotes the wheelbase, δ indicates the front wheel steering angle, and $\delta = \delta_{\mathrm{SW}}/i_{\mathrm{SW}}$ and i_{SW} are the total gear ratios of the steering system.

This study investigates ego vehicle trajectory prediction, with the core objective being to accurately infer trajectories within a future time window based on historical coupling features. The trajectory for the next T_f time windows is defined as the output feature matrix $\boldsymbol{Y} = [\boldsymbol{y}_{t+1}; \cdots; \boldsymbol{y}_{t+T_f}] \in \mathbb{R}^{T_f \times m}$, where $m = 2$, that is, only predicting the CG position coordinates $\hat{\boldsymbol{y}}_\tau = [\hat{x}_\tau, \hat{y}_\tau]$ of the vehicle in the global coordinate system.

In summary, the problem of ego vehicle trajectory prediction for human–machine co-driving vehicles in diverse urban scenarios can be formalized as follows: Based on the "driver-vehicle-road" coupling feature matrix $\boldsymbol{X}$ within the historical time window T_p, establish the mapping relation $f\colon \boldsymbol{X} \mapsto \boldsymbol{Y}$, i.e., solve the conditional probability distribution $p(\boldsymbol{Y}|\boldsymbol{X})$ to achieve precise prediction of the ego vehicle's position for the next T_f steps. The loss function employs the mean squared error (MSE) to ensure minimal positional deviation between the predicted and actual trajectories:

$$
L_{\mathrm{loss}} = \frac{1}{T_f} \sum_{\tau=t+1}^{t+T_f} \left|\left|\hat{\boldsymbol{y}}_\tau - \boldsymbol{y}_\tau\right|\right|_2^2 \tag{2}
$$

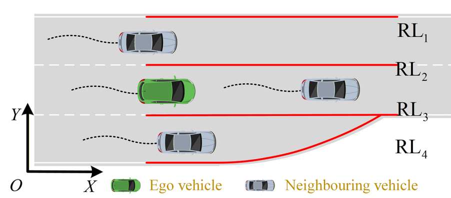

Fig. 2 Schematic of road geometric features. Red lines represent road coordinates input $\boldsymbol{s}^r$.

清華大学出版社 Tsinghua University Press | SciOpen

where $\boldsymbol{y}_\tau = [x_\tau, y_\tau]$ denotes the actual position at moment τ.

3 Trajectory prediction based on driver-vehicle coupling feature encoding

The proposed DriVETP framework centers on the cross-modal coupling features of "driver-vehicle-road", employing a three-tier architecture of "bidirectional temporal encoding of coupling features, cross-modal attention fusion, and temporal encoding", to achieve ego vehicle trajectory prediction. Its core logic lies in capturing temporal dependencies through BiLSTM while leveraging attention to focus on cross-modal correlations, thereby achieving precise mapping from the input feature matrix to the output trajectory sequence. Compared to Transformers, LSTMs offer higher efficiency for moderate sequence lengths and inherent compatibility with recurrent decoding. Attention mechanisms achieve dynamic feature weight allocation while avoiding the quadratic computational cost of global self-attention.

3.1 Bidirectional temporal encoding of coupling features

When a driver manipulates a vehicle, its current state relates not only to previous states but also to the driver's anticipated judgment of the future driving environment. Therefore, the input historical feature matrix exhibits bidirectional temporal dependencies. BiLSTM achieves bidirectional aggregation of contextual information by constructing forward and backward LSTM processing streams. This architecture enables the network to simultaneously capture past and future sequence features, demonstrating superior performance over unidirectional LSTMs in tasks such as sequence labeling and trajectory prediction (Qian et al., 2025).

Forward propagation generates the forward hidden state sequence from time step $t - T_p + 1$ to t, capturing the influence of past features on the current state:

$$\overrightarrow{\boldsymbol{h}}_\tau = \mathrm{LSTM}\left(\boldsymbol{h}^{\mathrm{pre}}_{\tau-1}, \overrightarrow{\boldsymbol{h}}_{\tau-1}\right) \tag{3}$$

where $\boldsymbol{h}^{\mathrm{pre}}_{\tau-1} \in \mathbb{R}^n$ denotes the preprocessed features at time step $\tau - 1$, $\overrightarrow{\boldsymbol{h}}_{\tau-1} \in \mathbb{R}^d$ represents the hidden state at the previous time step in the forward LSTM, and d indicates the dimension of the hidden layer.

Backpropagation generates a sequence of hidden states arranged in reverse chronological order from time step t to $t - T_p + 1$, thereby capturing the relationship between future features and the current state:

$$\overleftarrow{\boldsymbol{h}}_\tau = \mathrm{LSTM}\left(\boldsymbol{h}^{\mathrm{pre}}_{\tau+1}, \overleftarrow{\boldsymbol{h}}_{\tau-1}\right) \tag{4}$$

where $\boldsymbol{h}^{\mathrm{pre}}_{\tau+1} \in \mathbb{R}^n$ denotes the preprocessed features at time step $\tau + 1$, and $\overleftarrow{\boldsymbol{h}}_{\tau+1} \in \mathbb{R}^d$ represents the backward LSTM hidden state from the previous time step.

The BiLSTM concatenates the forward and backward hidden states at each time step to generate the encoding output, ensuring that features at each time step incorporate bidirectional temporal information, i.e., $\boldsymbol{h}^{\mathrm{enc}}_\tau = \left[\overrightarrow{\boldsymbol{h}}^T_\tau, \overleftarrow{\boldsymbol{h}}^T_\tau\right] \in \mathbb{R}^{2d}$. This ultimately yields the encoding sequence $\boldsymbol{H}^{\mathrm{enc}} = \left[\boldsymbol{h}^{\mathrm{enc}}_{t-T_p+1}; \cdots; \boldsymbol{h}^{\mathrm{enc}}_t\right] \in \mathbb{R}^{T_p \times 2d}$.

3.2 Cross-modal attention fusion

In the encoding sequence $\boldsymbol{H}^{\mathrm{enc}}$, the contribution of different modal features to trajectory prediction dynamically varies with operating conditions. A fixed weighting mechanism cannot adapt to this dynamic change. Therefore, we employ a multihead attention mechanism to dynamically allocate cross-modal weights according to operational requirements (Giuliari et al., 2021).

Map the encoding sequence $\boldsymbol{H}^{\mathrm{enc}}$ to the query space $\boldsymbol{Q}$, key space $\boldsymbol{K}$, and value space $\boldsymbol{V}$ as

$$\begin{cases} \boldsymbol{Q} = \boldsymbol{H}^{\mathrm{enc}} \boldsymbol{W}^Q \\ \boldsymbol{K} = \boldsymbol{H}^{\mathrm{enc}} \boldsymbol{W}^K \\ \boldsymbol{V} = \boldsymbol{H}^{\mathrm{enc}} \boldsymbol{W}^V \end{cases} \tag{5}$$

Among these, $\boldsymbol{W}^Q$, $\boldsymbol{W}^K$, and $\boldsymbol{W}^V$ are learnable parameters, while $\boldsymbol{Q}$, $\boldsymbol{K}$, $\boldsymbol{V} \in \mathbb{R}^{T_p \times \frac{2d}{h}}$, and h represents the number of attention heads, ensuring that the feature dimension can be uniformly partitioned into h parts.

Subsequently, feature weights are computed using the scalar dot product attention mechanism.

$$\mathrm{Attention}\left(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}\right) = \mathrm{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^T}{\sqrt{d_k}}\right)\boldsymbol{V} \tag{6}$$

where d_k represents the dimensions of $\boldsymbol{Q}$ and $\boldsymbol{K}$, i.e., $d_k = \frac{2d}{h}$, and $\sqrt{d_k}$ serves as the scaling factor to adjust the range of the dot product values. This prevents excessively large dot products from causing the gradients of the Softmax function to vanish or decay, thereby avoiding computational inefficiency.

Finally, the outputs from the h attention heads are concatenated columnwise. This concatenated output is then integrated through a fully connected layer to form the context vector $\boldsymbol{C} \in \mathbb{R}^{T_p \times 2d}$.

$$\begin{aligned} \boldsymbol{C} &= \mathrm{Concat}\left(\mathrm{head}_1, \ldots, \mathrm{head}_h\right) \cdot \boldsymbol{W}^O + \boldsymbol{b}^O \\ \mathrm{head}_i &= \mathrm{Attention}\left(\boldsymbol{Q}\boldsymbol{W}^Q_i, \boldsymbol{K}\boldsymbol{W}^K_i, \boldsymbol{V}\boldsymbol{W}^V_i\right) \end{aligned} \tag{7}$$

where $\boldsymbol{W}^Q_i$, $\boldsymbol{W}^K_i$, and $\boldsymbol{W}^V_i$ denote the projection matrices for each subspace, while $\boldsymbol{W}^O \in \mathbb{R}^{2d \times 2d}$ and $\boldsymbol{b}^O \in \mathbb{R}^{2d}$ represent learnable parameters used to integrate the output features. This architecture enables the model to capture complementary information from input features across different semantic spaces, effectively enhancing its ability to model complex spatiotemporal dependencies.

3.3 LSTM temporal encoding

The core objective of the decoding layer is to recursively generate future trajectory sequences based on the context vector $\boldsymbol{C}$. An LSTM is employed instead of a BiLSTM because the decoding process must generate sequences in chronological order, requiring only historical predictions and current context information—without needing reverse temporal dependencies.

The initial hidden state $\boldsymbol{s}_0$ and cell state $\boldsymbol{c}_0$ of the decoder layer LSTM are generated from the temporal mean of the context vector $\boldsymbol{C}$, ensuring that the initial state incorporates core information from all modalities (Yang et al., 2022).

$$\begin{aligned} \boldsymbol{s}_0 &= \tanh\left(\boldsymbol{W}_s \cdot \mathrm{Mean}\left(\boldsymbol{C}\right) + \boldsymbol{b}_s\right) \\ \boldsymbol{c}_0 &= \tanh\left(\boldsymbol{W}_c \cdot \mathrm{Mean}\left(\boldsymbol{C}\right) + \boldsymbol{b}_c\right) \end{aligned} \tag{8}$$

where $\mathrm{Mean}\left(\boldsymbol{C}\right)$ denotes the mean of context vector $\boldsymbol{C}$ across the temporal dimension, while $\boldsymbol{W}_s$, $\boldsymbol{W}_c \in \mathbb{R}^{2d \times d}$, and $\boldsymbol{b}_s$, $\boldsymbol{b}_c \in \mathbb{R}^d$ represent learnable parameters. The dimension of the LSTM hidden layer matches that of the encoding layer.

For the future time step $\tau = t + 1$ to $t + T_f$, the decoder input $\boldsymbol{y}^{\mathrm{dec}}_\tau = \left[\hat{\boldsymbol{y}}^T_{\tau-1}, \boldsymbol{c}^T_\tau\right]^T \in \mathbb{R}^{2d+2}$ is obtained by concatenating the previous prediction $\hat{\boldsymbol{y}}_{\tau-1} \in \mathbb{R}^2$ with the current context vector $\boldsymbol{c}_\tau \in \mathbb{R}^{2d}$. Subsequently, based on the input $\boldsymbol{y}^{\mathrm{dec}}_\tau$ and the previous state, the hidden state and cell state of the LSTM are updated as

$$\boldsymbol{s}_\tau, \boldsymbol{c}_\tau = \mathrm{LSTM}\left(\boldsymbol{y}_\tau^{\mathrm{dec}}, (\boldsymbol{s}_{\tau-1}, \boldsymbol{c}_{\tau-1})\right) \tag{9}$$

Finally, the future CG position predictions are generated through the fully connected layer.

$$\hat{\boldsymbol{y}}_\tau = \boldsymbol{W}_y \cdot \boldsymbol{s}_\tau + \boldsymbol{b}_y \tag{10}$$

where $\boldsymbol{W}_y \in \mathbb{R}^{d\times 2}$ and $\boldsymbol{b}_y \in \mathbb{R}^2$ are learnable parameters.

4 Experimental validation and results analysis

We validate DriVETP's prediction accuracy and generalization capability by collecting driving data from nonprofessional drivers in urban scenarios using a hardware-in-the-loop dynamic driving simulator. The model is trained and validated through common urban driving scenarios such as lane changing, intersection turning, and roundabout navigation, with its prediction accuracy compared against mainstream algorithms. We analyzed DriVETP's predictive performance in diverse urban scenarios driven by driver-vehicle coupling features through multidimensional metrics, including prediction accuracy, stability, and computational efficiency, and explored its generalization capabilities.

4.1 Data collection and processing

The driving data used to validate the prediction algorithm in this study are collected via a dynamic driving simulator, as shown in Fig. 3. This simulator comprises five core components: the NI PXIe-1088, Speedgoat (Intel Core i7 4.2 GHz), the driving system, and two host PCs. Speedgoat serves as the system hub,

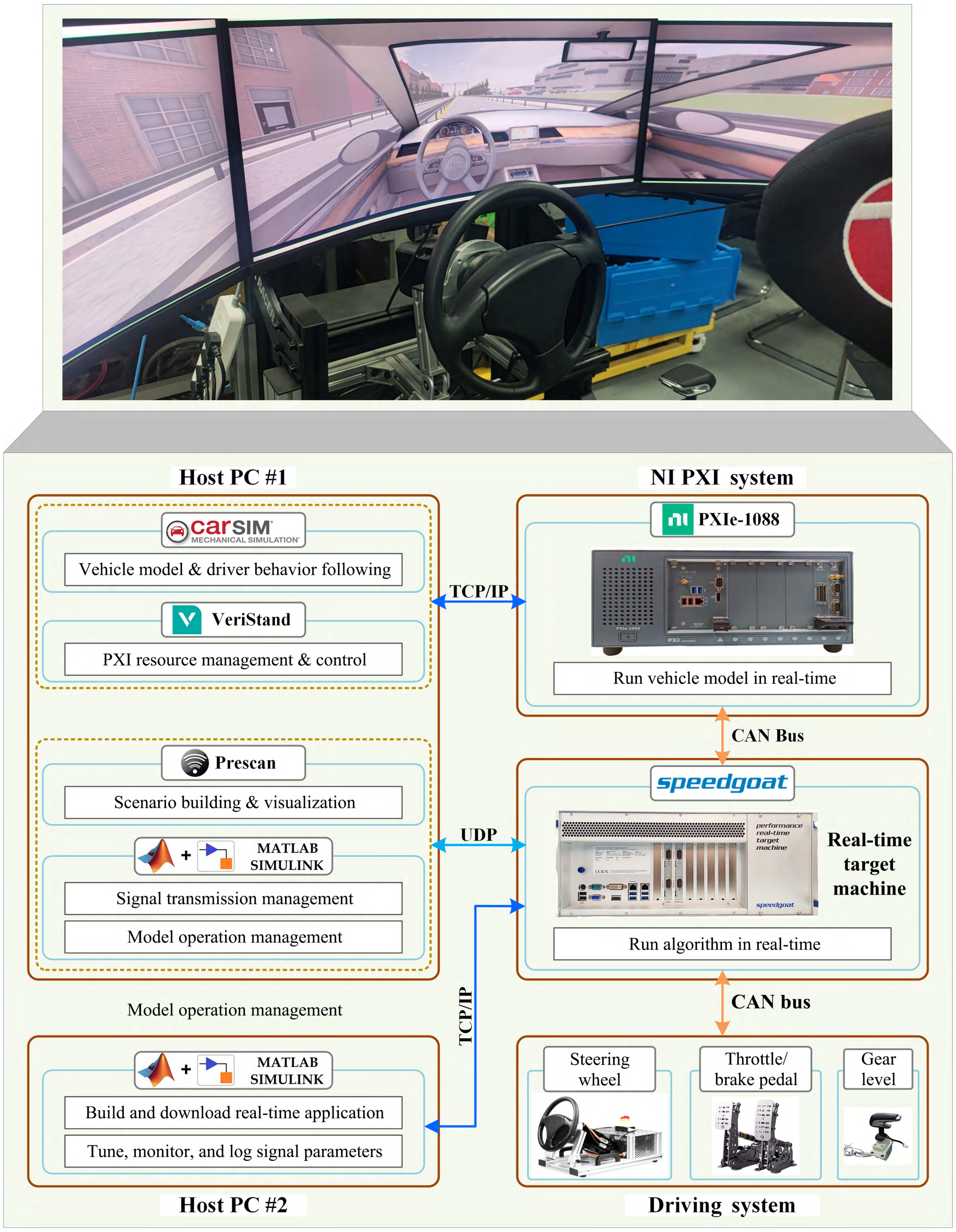

Fig. 3 Dynamic driving simulator.

transmitting driving system control commands to the vehicle model within the NI PXIe-1088 via CAN bus while receiving status feedback from the vehicle model. It simultaneously transmits vehicle status to host PC #1 via UDP. Prescan 8.5.0 running on Host PC #1 renders an urban scene featuring intersections, roundabouts, and continuous curves, receiving vehicle status to synchronize driving behavior within the simulation environment. Host PC #1 concurrently runs both VeriStand 2023 Q1 and CarSim 2021.0, communicating with the NI PXIe-1088 via TCP/IP. CarSim configures the vehicle dynamics model, while VeriStand configures the NI PXIe-1088's I/O channels and user interface. Host PC #2 connects to the Speedgoat via TCP/IP for compiling, downloading, and debugging models running within Speedgoat. The steering system employs a Sensodrive Force-Feedback Steering Wheel that simulates authentic road feel. Vehicle dynamics are controlled via the brake and accelerator pedals, with the control loop operating at 100 Hz and minimal manual latency (< 10 ms end-to-end). These features ensure that the driver-vehicle interaction closely replicates real vehicle response characteristics.

The scenes and driving routes constructed in Prescan are shown in Fig. 4. The red line indicates the actual driving route, while the green dot marks the route's starting point. The figure also displays enlarged views of several typical road scenarios. Two nonprofessional drivers performed random driving maneuvers in the simulator. Their control inputs and vehicle states were captured in real time and combined with future road geometric features to form the dataset. This dataset contains extensive driving data, including lane changing, intersection turning, and roundabout navigation data. One driver was selected as the primary research subject, generating approximately 600 min of driving data in the simulator. The other driver's 40 min of driving data were used to validate the algorithm's generalization capability.

In real-world driving scenarios, nonprofessional drivers exhibit poorer driving skills, making it difficult to achieve repeatable driving behaviors and vehicle states even under identical conditions (Wei et al., 2025). This characteristic provides ideal conditions for testing the generalization capabilities of predictive models. By selecting driving data from nonprofessional drivers, we aim to introduce highly variable motion trajectories that simulate the diverse driving behaviors found in the real world, thereby increasing the challenge of trajectory prediction. Compared to professional drivers with more consistent behaviors, the random operations of nonprofessional drivers force neural networks to learn more generalizable prediction patterns, thereby avoiding overfitting to average driving behaviors (Paparusso et al., 2023).

The initial sampling frequency of the dynamic driving simulator is 100 Hz. To balance data detail and computational efficiency, we downsample the data to a 4 Hz frequency. To fully utilize the collected driving data, we extract training samples through a sliding time window. At time step t, we selected 23-dimensional source data from the time interval $t-T_p+1: t$, encompassing spatial motion features, driver-vehicle coupling features, and road geometric features. Target data comprise 2-dimensional vehicle CG position features from time steps $t+1: t+T_f$. Here, $T_p=24$ and $T_f=20$, meaning we predict the target trajectory for the next 5 s using source data from the preceding 6 s. This process yielded a total of 101,370 sample pairs. The samples were then split into training (70,959 pairs), validation (20,274 pairs), and test (10,137 pairs) sets at a 7:2:1 ratio. The number of typical scenarios in each sample is shown in Table 1. Subsequently, the data were standardized using Z-scores to conform to a standard normal distribution, eliminating the impact of different dimensions or data distribution differences on model training.

4.2 Analysis of model performance

To validate DriVETP's predictive performance in diverse urban scenarios, we design multiple comparative experiments. AM-LSTM (Wang et al., 2024) and Transformer (Giuliari et al., 2021) serve as baseline models, demonstrating the performance gap between DriVETP and mainstream algorithms. DriVETP-Ablation1 and DriVETP-Ablation2 served as ablation experiments. The former removes driver manipulation features from input features, validating their core role in trajectory prediction. The latter replaces BiLSTM with LSTM to validate the advantage of BiLSTM in capturing temporal dependencies. The DriVETP and ablation experiments share the same set of model parameters, which can be obtained from Fig. 1.

Common evaluation metrics in trajectory prediction research include root mean squared error (RMSE), average displacement error (ADE), final displacement error (FDE), and computation time. The RMSE clearly quantifies the discrepancy between the predicted and actual values, but its stability is compromised by outliers. ADE provides a holistic measure of the average distance error between predictions and reality, yet it focuses solely on average error and remains insensitive to the presence of outliers. FDE reflects local errors at individual time points but fails to capture overall algorithmic performance. The computation time

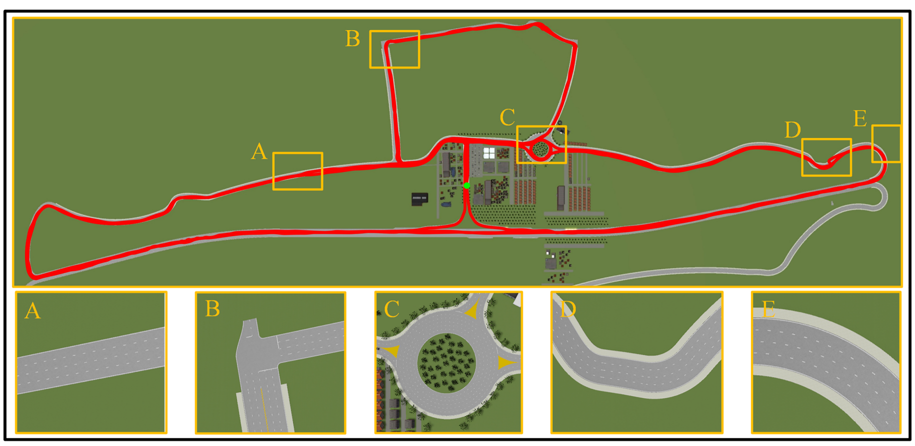

Fig. 4 Panoramic view of the data collection scenario and driving route. The red line indicates the driving route, while the green dot marks the starting point. A–E show enlarged views of specific areas.

Table 1 Number of typical scenarios in the dataset

Case	Number
Left lane-changing	4004
Right lane-changing	3665
Roundabout navigation	2849
Left turning	2187
Right turning	4901

solely reflects the computational efficiency. This study aims to achieve real-time, high-precision trajectory prediction in diverse urban scenarios. To overcome the limitations of single-metric evaluations and comprehensively compare model performance, we propose the Comprehensive Performance Evaluation (CPE) metric (Zhang et al., 2024b), as shown in Eq. (11). This metric measures accuracy through the mean and stability through variance. By simultaneously calculating the mean and variance of RMSE, ADE, and FDE alongside computation time, CPE achieves a comprehensive quantitative assessment of prediction accuracy, stability, and computational efficiency, balancing overall and local errors. A lower CPE value indicates superior model performance.

$$\begin{aligned}\mathrm{CPE}\,(i) = & \frac{\mathrm{RMSE\_}M\,(i)}{\max\,(\mathrm{i}tRMSE\_M\,(i))} + \frac{\mathrm{RMSE\_}V\,(i)}{\max\,(\mathrm{RMSE\_}V\,(i))} \\ & + \frac{\mathrm{ADE\_M}\,(i)}{\max\,(\mathrm{ADE\_}M\,(i))} + \frac{\mathrm{ADE\_}V\,(i)}{\max\,(\mathrm{ADE\_}V\,(i))} \\ & + \frac{\mathrm{FDE\_}M\,(i)}{\max\,(\mathrm{FDE\_}M\,(i))} + \frac{\mathrm{FDE\_}V\,(i)}{\max\,(\mathrm{FDE\_}V\,(i))} \\ & + \frac{T\_M\,(i)}{\max\,(T\_M\,(i))}\end{aligned} \tag{11}$$

where $i =$ DriVETP, DriVETP-Ablation1, DriVETP-Ablation2, AM-LSTM, and Transformer. $*\_M$ and $*\_V$ denote the mean and variance of $*$, respectively, and T_M denotes the average computation time.

Different models exhibit significant variations across six typical urban scenario test cases, as illustrated in Fig. 5. Among them, DriVETP, DriVETP-Ablation2, AM-LSTM, and Transformer share identical input features encompassing spatial kinematic features, driver-vehicle coupling features, and road geometry features. DriVETP-Ablation1, however, excludes driver manipulation features from its input. As shown in Fig. 5, incorporating driver manipulation features during model training enables thorough learning of driver behavior during steering maneuvers. This results in prediction curves that more closely align with actual outcomes in scenarios such as lane changing, turning, and roundabout navigation. The RMSE for the predicted trajectories around the roundabout is presented in Table 2. Compared to DriVETP-Ablation1, the remaining four models all demonstrate superior prediction performance. DriVETP-Ablation2 and DriVETP exhibit nearly identical RMSE values. However, as the prediction time horizon extends, DriVETP-Ablation2 demonstrates more pronounced trajectory prediction deviation fluctuations.

During steering transitions (e.g., lane changing or turning initiation), DriVETP also demonstrates superior alignment performance. At these moments, human-vehicle coupling intensity peaks, indicating that the coupling features effectively capture these dynamic characteristics. Relatively larger errors occur during straight driving, a phenomenon consistent with expectations. This scenario exhibits simple dynamic characteristics where baseline models perform well; the slight increase in error may reflect model overdesign for static conditions. Additionally, Table 2 provides prediction errors for the state-of-the-art (SOTA) models GaVA (Liao et al., 2024) and DEMO (Wang et al., 2025b) in specific scenarios over the past two years. It can be observed that while DriVETP does not demonstrate superior prediction accuracy to SOTA within the 5-second prediction horizon, its prediction performance remains highly competitive compared to mainstream approaches relying on multivehicle interaction information. This advantage stems from its encoding mechanism based on driver-vehicle coupling features. By explicitly encoding coupling features with clear physical significance, the model not only enables intuitive observation of prioritized dynamic relationships across different operational scenarios through attention visualization techniques but also achieves more stable prediction outcomes. Furthermore, due to its lightweight architecture, it also shows competitive computational efficiency (< 20 ms inference).

Fig. 5 also shows that DriVETP's predicted values align most closely with actual values. This demonstrates the effectiveness of DriVETP's bidirectional temporal dependency encoding in trajectory prediction features learning. Compared to AM-LSTM and Transformer, DriVETP also delivers superior performance. Since both algorithms employ attention mechanisms for global feature encoding, their predictive capabilities lack significant competitiveness without specific embedding methods or encoding patterns. Furthermore, their modeling performance for short-to-medium-term time series falls short of LSTM's excellence.

Fig. 6 compares the mean and variance of performance metrics across six typical scenarios, revealing the average computational efficiency of different models. As shown, in terms of prediction accuracy, the DriVETP model—which incorporates driver-vehicle coupling characteristics—exhibits the closest mean and variance to the center of the radar chart across all scenarios. Its highly consistent prediction accuracy across different scenarios means that DriVETP's radar chart essentially forms a regular hexagon. This indicates that the model delivers precise and stable predictions in diverse urban scenarios.

Table 3 presents multidimensional quantitative metrics of prediction performance across six typical scenarios for different models. Regarding prediction accuracy, DriVETP achieved a 55.57% improvement in RMSE_M compared to DriVETP-Ablation1, a –4.99% improvement over DriVETP-Ablation2, a 42.46% improvement versus AM-LSTM, and a 51.61% improvement relative to Transformer. Compared to the DriVETP-Ablation1 model that disregards driver manipulation achieved 57.15% and 59.28% improvements in ADE_M and FDE_M, respectively. Against the LSTM-encoded DriVETP-Ablation2, it recorded −1.8% and 30.58% improvements, while against AM-LSTM and Transformer, DriVETP also achieved 40% improvements in ADE_M and FDE_M. Regarding prediction stability, DriVETP's RMSE_V outperformed the other algorithms by 52.93%, −1.26%, 48.85%, and 57.49%, respectively, while ADE_V improved by 53.47%, –4.59%, 52.90%, and 61.52%, respectively. FDE_V improved by 63.11%, 60.00%, 55.43%, and 65.60%, respectively. This outcome aligns with previous analyses: although DriVETP-Ablation2 with LSTM achieved superior overall error performance compared to DriVETP, its stability fell significantly short of DriVETP employing BiLSTM for feature encoding.

In terms of computational efficiency, we analyzed the overall computation time of five models predicting the future time horizon, as shown in Fig. 7. All experiments were conducted on a

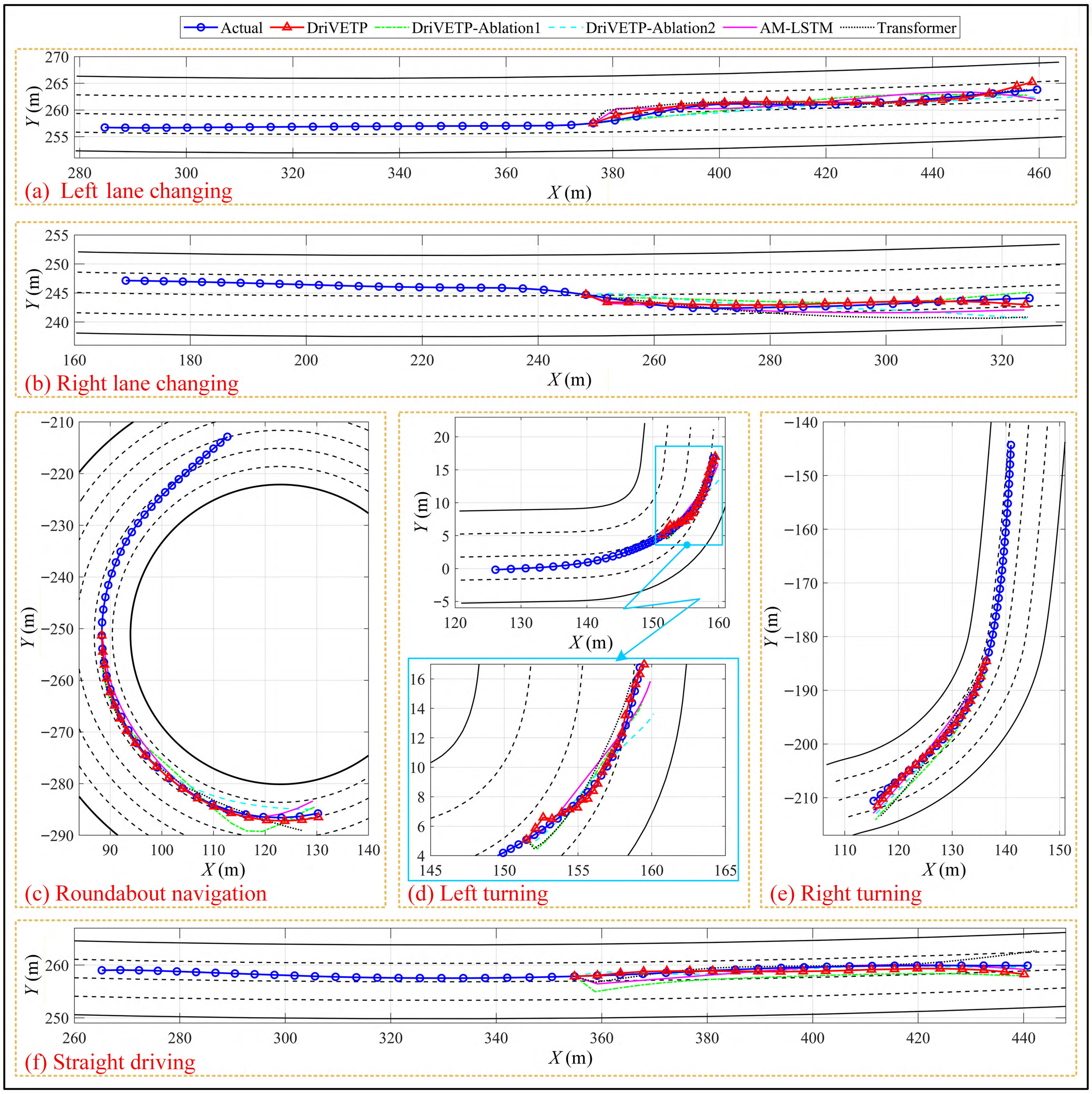

Fig. 5 Comparison of ego vehicle trajectory prediction results across different models in diverse urban scenarios. (a) Test scenarios include left lane changing, (b) right lane changing, (c) roundabout navigation, (d) left turning, (e) right turning, and (f) straight driving.

Table 2 RMSE over a 5 s prediction horizon for different prediction models in the roundabout case (Unit: m)

Method	Prediction horizon				
	1 s	2 s	3 s	4 s	5 s
DriVETP	1.02	0.81	0.70	0.61	0.62
DriVETP-Ablation1	2.40	2.42	2.05	1.93	2.03
DriVETP-Ablation2	0.69	0.50	0.62	0.71	1.01
AM-LSTM	1.74	1.34	1.39	1.29	1.21
Transformer	2.14	1.65	1.41	1.29	1.21
GaVA	0.17	0.24	0.42	0.80	1.31
DEMO	0.06	0.14	0.25	0.44	0.70

computer equipped with an Intel Core i7-10750H CPU @2.60 GHz, NVIDIA GeForce GTX 1650Ti 4 GB GPU, and 32 GB RAM. Although the figure indicates minimal differences in overall computation time across models, with times consistently below 20 ms, this performance fully meets the 100 ms runtime requirement for motion planning. It outperforms the ego vehicle trajectory prediction algorithm described in the paper. Furthermore, since onboard computing devices significantly

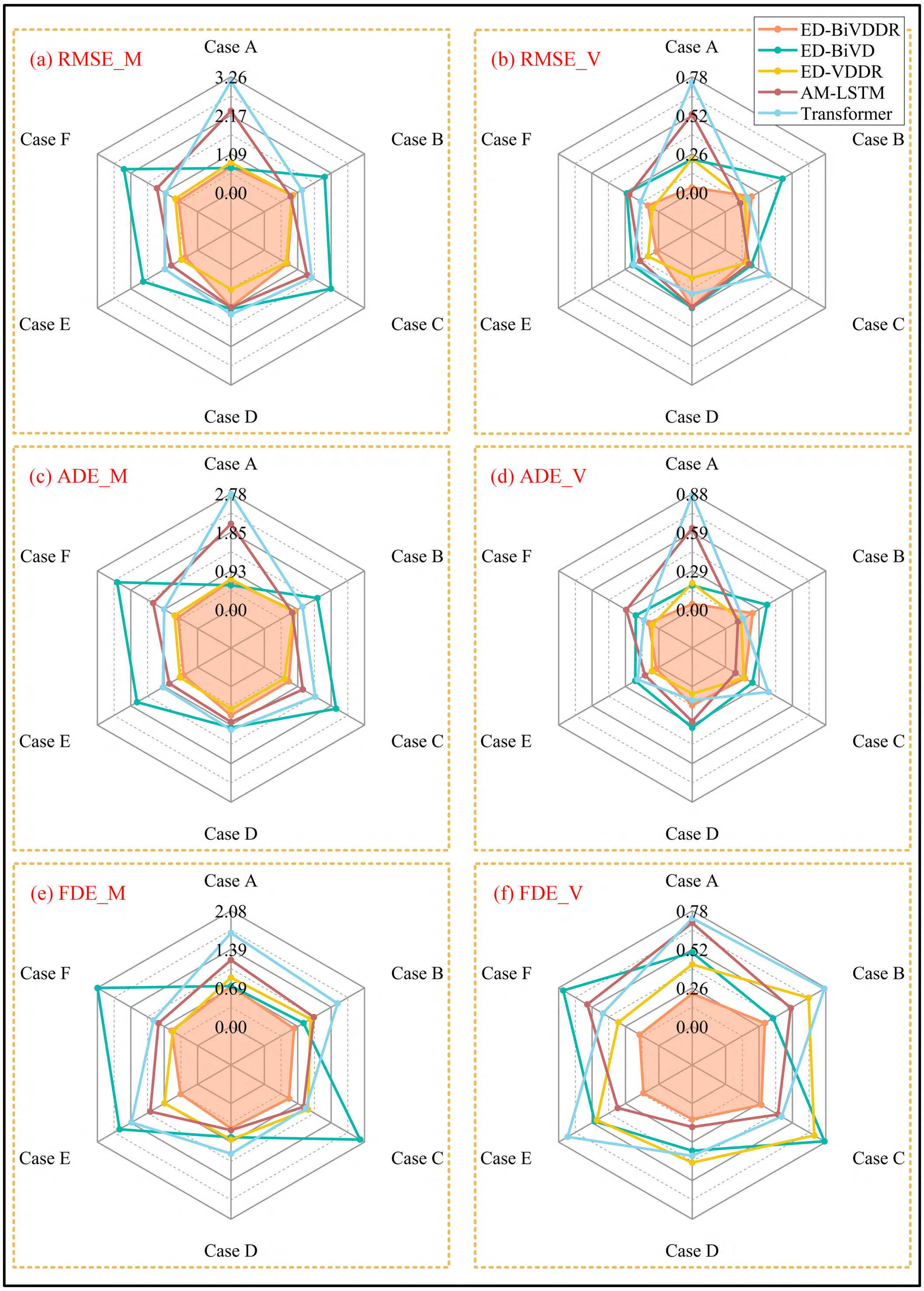

Fig. 6 Radar chart of average performance metrics for 6 cases in urban scenarios.

Table 3 Evaluation and comparison of experimental results for various cases in urban scenarios

Method	Evaluation index							CPE
	RMSE_M (m)	RMSE_V (m)	ADE_M (m)	ADE_V (m)	FDE_M (m)	FDE_V (m)	T_M (s)	
DriVETP	0.7474	0.1240	0.6388	0.1167	0.5316	0.1968	0.0129	3.0631
DriVETP-Ablation1	$1.6824_{\downarrow 55.57\%}$	$0.2633_{\downarrow 57.15\%}$	$1.4906_{\downarrow 59.28\%}$	$0.2508_{\downarrow 52.93\%}$	$1.3054_{\downarrow 53.47\%}$	$0.5334_{\downarrow 63.11\%}$	0.0120	$6.2476_{\downarrow 50.97\%}$
DriVETP-Ablation2	$0.7119_{\uparrow 4.99\%}$	$0.1224_{\uparrow 1.80\%}$	$0.6275_{\uparrow 30.58\%}$	$0.1116_{\uparrow 1.25\%}$	$0.7658_{\uparrow 4.59\%}$	$0.4919_{\downarrow 60.00\%}$	0.0096	$3.5466_{\downarrow 13.63\%}$
AM-LSTM	$1.2989_{\downarrow 42.46\%}$	$0.2424_{\downarrow 43.38\%}$	$1.1281_{\downarrow 39.93\%}$	$0.2478_{\downarrow 48.86\%}$	$0.8850_{\downarrow 52.91\%}$	$0.4415_{\downarrow 55.43\%}$	0.0079	$5.0120_{\downarrow 38.88\%}$
Transformer	$1.5445_{\downarrow 51.61\%}$	$0.2916_{\downarrow 53.06\%}$	$1.3610_{\downarrow 56.02\%}$	$0.3032_{\downarrow 57.50\%}$	$1.2087_{\downarrow 61.52\%}$	$0.5721_{\downarrow 65.60\%}$	0.0205	$6.7570_{\downarrow 54.67\%}$

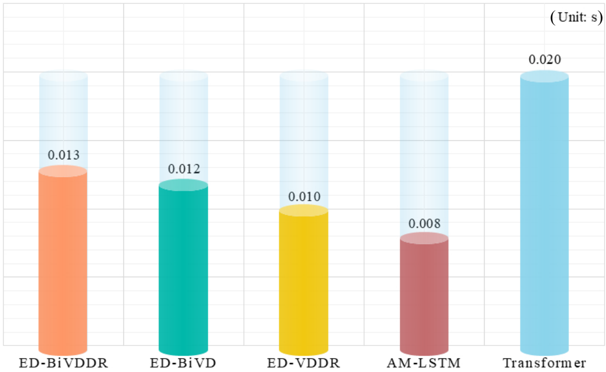

Fig. 7 Bar chart showing the average calculation time (s) for different models in diverse urban scenarios.

exceed current laptop capabilities, this computational efficiency will be further enhanced in practical applications.

Additionally, the CPE metrics in Table 3 indicate that DriVETP demonstrates the best overall performance across multiple models, achieving improvements of 50.97%, 13.58%, 38.86%, and 54.65% over DriVETP-Ablation1, DriVETP-Ablation2, AM-LSTM, and Transformer, respectively. In terms of CPE, DriVETP achieves a 12.58% improvement over DriVETP-Ablation2. The distinction between the two models lies in DriVETP's ability to capture richer temporal dependencies through BiLSTM encoding compared to LSTM encoding. The roundabout navigation prediction results in Table 2 also demonstrate comparable prediction performance between the two models, although DriVETP exhibits lower local error. Compared to DriVETP-Ablation1, DriVETP incorporates driver manipulation behaviors into its input features. By encoding driver-vehicle coupling features that include these manipulations, the prediction model gains dynamic associations between driver manipulations and vehicle states, yielding results closer to human driving scenarios. The lowest CPE achieved by DriVETP relative to the other models further validates the effectiveness of the proposed framework.

4.3 Generalization verification

The above analysis validates the effectiveness of enhancing DriVETP's ego vehicle trajectory prediction performance in diverse urban scenarios by considering driver manipulation characteristics, which is the core objective of this work. However, unlike autonomous vehicles, human–machine co-driving vehicles involve different drivers. Therefore, ensuring that the ego vehicle prediction algorithm achieves high prediction accuracy under the control of different drivers is a critical consideration for the algorithm. To address this, we collected an additional 40 min of driving data from another driver during data acquisition. From this dataset, we selected scenarios involving left and right lane changing for generalization validation.

Fig. 8 displays DriVETP's predicted trajectories for lane changing under another driver's manipulation, with the corresponding RMSE listed in Table 4. The predicted trajectories in case (a) and case (b) closely match the actual trajectories, indicating that the algorithm maintains high prediction accuracy even when applied to vehicles under manipulation by other drivers. Cases (a) and (b) represent lane-changing scenarios, where the prediction accuracy is slightly lower than that in the previous results. Nevertheless, the RMSE remains within 1.5 m over a 5-s prediction horizon.

We acknowledge that these results, while promising, are based on simulator data from only two drivers; validation with larger, real-world driver populations remains essential future work.

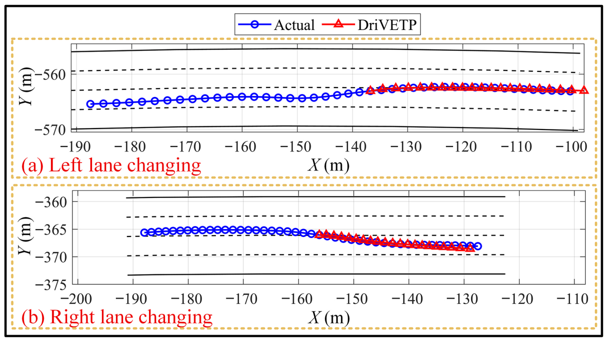

Fig. 8 Trajectory prediction based on another driver's driving data.

Table 4 RMSE over a 5 s prediction horizon for another driver in different urban scenarios (Unit: m)

	Prediction horizon					
	1 s	2 s	3 s	4 s	5 s	
Case (a)	0.68	0.79	1.03	1.07	1.06	< 1.5 m
Case (b)	0.43	0.44	0.65	0.92	1.17	< 1.5 m

5 Conclusions

This study proposes DriVETP, a deep learning trajectory prediction framework based on driver-vehicle coupling feature encoding. It aims to predict the deterministic short-to-medium-term driving trajectory of the ego vehicle under driver manipulation in diverse urban scenarios. The framework constructs a three-dimensional coupled feature system encompassing "driver-vehicle-road", leveraging BiLSTM to capture bidirectional temporal dependencies and employing multihead attention mechanisms to dynamically focus on cross-modal key correlations. Ultimately, an LSTM recursively generates precise trajectories. Using urban driving data collected from the dynamic driving simulator, DriVETP's predictive performance is validated through comparisons with mainstream models and ablation experiments. Experimental results demonstrate DriVETP' s outstanding prediction accuracy and stability across diverse urban scenarios—including lane changing, turning, and roundabout navigation—while delivering exceptional real-time forecasting capabilities. This framework effectively addresses existing research limitations such as biased feature capture, insufficient real-time performance, and weak generalization. It provides precise trajectory inputs for risk assessment and safety decision-making in human–machine co-driving systems.

We acknowledge that comprehensive scenario diversity extends beyond the typical maneuvers examined in this study to encompass more complex interaction scenarios, such as unprotected left turns, dense pedestrian zones, and lane merging during peak hours. In future work, we will collect data and conduct training for these scenarios to further enhance the model' s adaptability across diverse environments.

Future research directions can be further expanded. First, personalized driver characteristics can be incorporated to build adaptive prediction models reflecting individual driving variations. Second, models are validated and optimized using large-scale real-world driving data while enriching algorithms with more comprehensive human-vehicle interaction information. Third, deeply integrate prediction results into downstream decision-making and motion planning modules of intelligent systems to achieve a closed-loop unified prediction-decision-planning process to achieve more proactive human–machine collaboration.

Author contributions

Guoying Chen: Funding acquisition, Conceptualization, Supervision, Writing – review & editing. **Zheng Gao:** Writing – original draft, Methodology, Software, Formal analysis, Visualization. **Tianjun Sun:** Funding acquisition, Conceptualization, Supervision, Writing – review & editing. **Xinyu Wang:** Writing – review & editing, Formal Analysis, Methodology. **Min Hua:** Investigation, Resources, Writing – review & editing. **Zhenhai Gao:** Conceptualization, Methodology, Supervision.

Replication and data sharing

The codes necessary for replication of this study are available at https://doi.org/10.26599/ETSD.2026.9190026.

Acknowledgements

This work was supported in part by the Jilin Provincial Department of Science and Technology Youth Science and Technology Talent Cultivation Project (No. 20250602051RC), the National Natural Science Foundation of China (No. 52372412), and Fundamental Research Funds for the Central Universities (No. 2025- JCXK-19).

Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article.

References

Ansari, S., Naghdy, F., Du, H., 2022. Human-machine shared driving: Challenges and future directions. *IEEE Trans Intell Veh*, **7**, 499–519.

Bharilya, V., Kumar, N., 2024. Machine learning for autonomous vehicle's trajectory prediction: A comprehensive survey, challenges, and future research directions. *Veh Commun*, **46**, 100733.

Bikmukhametov, T., Jäschke, J., 2020. Combining machine learning and process engineering physics towards enhanced accuracy and explainability of data-driven models. *Comput Chem Eng*, **138**, 106834.

Camarcat, L., Feng, Y., Formosa, N., Quddus, M., 2025. Hierarchical Bayesian threshold excess model for real-time vehicle-based conflict prediction in dynamic traffic environments. *Commun Transp Res*, **5**, 100210.

Chen, G., Gao, Z., Hua, M., Shuai, B., Gao, Z., 2024. Lane change trajectory prediction considering driving style uncertainty for autonomous vehicles. *Mech Syst Signal Process*, **206**, 110854.

Fang, J., Wang, F., Xue, J., Chua, T. S., 2024. Behavioral intention prediction in driving scenes: A survey. *IEEE Trans Intell Transp Syst*, **25**, 8334–8355.

Fang, Z., Wang, J., Wang, Z., Liang, J., Liu, Y., Yin, G., 2023. A human-machine shared control framework considering time-varying driver characteristics. *IEEE Trans Intell Veh*, **8**, 3826–3838.

Gao, H., Qin, Y., Hu, C., Liu, Y., Li, K., 2023. An interacting multiple model for trajectory prediction of intelligent vehicles in typical road traffic scenario. *IEEE Trans Neural Netw Learn Syst*, **34**, 6468–6479.

Gao, Z., Jia, B., Xie, D., Wang, W., Wu, J., 2025. A discussion on the complexity and transit mechanisms of urban traffic systems. *Engineering*, **44**, 24–29.

Geng, M., Li, J., Xia, Y., Chen, X. M., 2023. A physics-informed Transformer model for vehicle trajectory prediction on highways. *Transp Res Part C Emerg Technol*, **154**, 104272.

Giuliari, F., Hasan, I., Cristani, M., Galasso, F., 2021. Transformer networks for trajectory forecasting. In: 2020 25th International Conference on Pattern Recognition (ICPR), 10335–10342.

Hu, H., Xiong, Z., Li, Z., Sun, T., Ran, R., 2026. Semi-supervised risk assessment research for intelligent vehicles inspired by collective biological risk-avoidance behaviors. *J Bionic Eng*, **23**, 225–238.

Hu, Y., Jia, X., Tomizuka, M., Zhan, W., 2022. Causal-based time series domain generalization for vehicle intention prediction. In: 2022 International Conference on Robotics and Automation (ICRA), 7806–7813.

Huang, T., Fu, R., Sun, Q., Deng, Z., Huang, S., Jin, L., 2025. Toward human-vehicle collaboration for automated vehicles: A review and perspective. *IEEE Trans Intell Transp Syst*, **26**, 5649–5673.

Huang, T., Fu, R., Sun, Q., Deng, Z., Liu, Z., Jin, L., et al., 2024. Driver lane change intention prediction based on topological graph constructed by driver behaviors and traffic context for human-machine co-driving system. *Transp Res Part C Emerg Technol*, **160**, 104497.

Huang, Y., Du, J., Yang, Z., Zhou, Z., Zhang, L., Chen, H., 2022. A survey on trajectory-prediction methods for autonomous driving. *IEEE Trans Intell Veh*, **7**, 652–674.

Hussain, F., Li, Y., Haque, S. M. M., 2025. Machine learning-based real-time crash risk forecasting for pedestrians. *Commun Transp Res*, **5**, 100224.

Jiang, Y., Zhu, B., Yang, S., Zhao, J., Deng, W., 2023. Vehicle trajectory prediction considering driver uncertainty and vehicle dynamics based on dynamic Bayesian network. *IEEE Trans Syst Man Cybern Syst*, **53**, 689–703.

Liao, H., Liu, S., Li, Y., Li, Z., Wang, C., Li, Y., et al., 2024. Human observation-inspired trajectory prediction for autonomous driving in mixed-autonomy traffic environments. In: 2024 IEEE International Conference on Robotics and Automation (ICRA), 14212–14219.

Liu, H. X., Feng, S., 2024. Curse of rarity for autonomous vehicles. *Nat Commun*, **15**, 4808.

Liu, S., Zheng, K., Zhao, L., Fan, P., 2020. A driving intention prediction method based on hidden Markov model for autonomous driving. *Comput Commun*, **157**, 143–149.

Liu, Y., Wu, F., Liu, Z., Wang, K., Wang, F., Qu, X., 2023. Can language models be used for real-world urban-delivery route optimization? *Innovation*, **4**, 100520.

Meng, Q., Guo, H., Liu, Y., Chen, H., Cao, D., 2023. Trajectory prediction for automated vehicles on roads with lanes partially covered by ice or snow. *IEEE Trans Veh Technol*, **72**, 6972–6986.

Messaoud, K., Cord, M., Alahi, A., 2025. Towards generalizable trajectory prediction using dual-level representation learning and adaptive prompting. In: 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 27564–27574.

Mozaffari, S., Al-Jarrah, O. Y., Dianati, M., Jennings, P., Mouzakitis, A., 2022. Deep learning-based vehicle behavior prediction for autonomous driving applications: A review. *IEEE Trans Intell Transp Syst*, **23**, 33–47.

Muther, T., Dahaghi, A. K., Syed, F. I., Van Pham, V., 2023. Physical laws meet machine intelligence: Current developments and future directions. *Artif Intell Rev*, **56**, 6947–7013.

Paparusso, L., Melzi, S., Braghin, F., 2023. Real-time forecasting of driver-vehicle dynamics on 3D roads: A deep-learning framework leveraging Bayesian optimisation. *Transp Res Part C Emerg Technol*, **156**, 104329.

Qian, H., Shi, X., Hawbani, A., Bi, Y., Liu, Z., Muthanna, A., et al., 2025. TRACER: Transfer knowledge-based collaborative vehicle trajectory prediction for highway traffic toward cross-region adaptivity. *IEEE Trans Intell Transp Syst*, **26**, 10747–10763.

Reisinger, S., Adelberger, D., del Re, L., 2021. A two-layer switching based trajectory prediction method. *Eur J Control*, **62**, 143–150.

Schuetz, E., Flohr, F. B., 2024. A review of trajectory prediction methods for the vulnerable road user. *Robotics*, **13**, 1.

Sun, T., Gao, Z., Chang, Z., Zhao, K., 2021. Brain-like intelligent decision-making based on basal ganglia and its application in automatic car-following. *J Bionic Eng*, **18**, 1439–1451.

Tan, Z., Dai, N., Su, Y., Zhang, R., Li, Y., Wu, D., et al., 2022. Human–machine interaction in intelligent and connected vehicles: A review of status quo, issues, and opportunities. *IEEE Trans Intell Transport Syst*, **23**, 13954–13975.

Wang, C., Liao, H., Li, Z., Xu, C., 2025a. WAKE: Towards robust and physically feasible trajectory prediction for autonomous vehicles with WAvelet and KinEmatics synergy. *IEEE Trans Pattern Anal Mach Intell*, **47**, 3126–3140.

Wang C., Liao H., Zhu K., Zhang G., Li Z., 2025b. DEMO: A Dynamics-Enhanced Learning Model for multi-horizon trajectory prediction in autonomous vehicles. *Inform Fusion*, **118**, 102924.

Wang, J., Liu, K., Li, H., 2024. LSTM-based graph attention network for vehicle trajectory prediction. *Comput Netw*, **248**, 110477.

Wei, C., Qin, Z., Li, S., Zhang, Z., Zhao, X., Abdelraouf, A., et al., 2025. PDB: Not All Drivers Are the Same—A Personalized Dataset for Understanding Driving Behavior. https://arxiv.org/abs/2503.06477

Xing, Y., Lv, C., Wang, H., Wang, H., Ai, Y., Cao, D., et al., 2019. Driver lane change intention inference for intelligent vehicles: Framework, survey, and challenges. *IEEE Trans Veh Technol*, **68**, 4377–4390.

Yang, Z., Liu, D., Ma, L., 2022. Vehicle trajectory prediction based on LSTM network. In: 2022 International Conference on Artificial Intelligence and Computer Information Technology (AICIT), 1–4.

Zhang, S., Zhang, J., Yang, L., Chen, F., Li, S., Gao, Z., 2024a. Physics guided deep learning-based model for short-term origin–destination demand prediction in urban rail transit systems under pandemic. *Engineering*, **41**, 276–296.

Zhang, Y., Zhou, Q., Wang, J., Kouvelas, A., Makridis, M. A., 2025. CASAformer: Congestion-aware sparse attention transformer for traffic speed prediction. *Commun Transp Res*, **5**, 100174.

Zhang, Z., Wang, C., Zhao, W., Cao, M., Liu, J., 2024b. Ego vehicle trajectory prediction based on time-feature encoding and physics-intention decoding. *IEEE Trans Intell Transp Syst*, **25**, 6527–6542.

Zhao, X., Chen, G., Gao, Z., Yao, J., Gao, Z., Hua, M., 2024. Autonomous obstacle avoidance for distributed drive electric vehicles via dynamic drifting. *IEEE Trans Transp Electrif*, **10**, 8893–8906.

Guoying Chen is currently a Professor with State Key Laboratory of Automotive Chassis Integration and Bionics at Jilin University, China, in 2012. He received the Ph.D. degree from Jilin University in 2012. His research interests include control-by-wire vehicle control, advanced vehicle control and autonomous vehicles.

Zheng Gao received the M.E. degree in vehicle engineering from Jilin University in 2022. He is currently working toward a Ph.D. degree with College of Automotive Engineering at Jilin University, China. His research interests include motion planning and motion control of intelligent vehicles.

Tianjun Sun received the Ph.D. degree in automotive engineering from Jilin University in 2019. His main research direction is new energy intelligent vehicle decision-making and control methods. Now he works as an Associate Professor at Jilin University. With the continuous deepening of research and multidisciplinary cross fusion, he tends to explore brain-like decision-making methods based on reinforcement learning and artificial intelligence.

Xinyu Wang received the B.E. degree in vehicle engineering from Jilin University in 2021. He is currently pursuing the Ph.D. degree in vehicle engineering with the College of Automotive Engineering, Jilin University, China. His research interests include autonomous driving and reinforcement learning and its applications in automotive systems.

Min Hua received the M.E. degree in vehicle engineering from Jilin University, China in 2012, and the Ph.D. degree in mechanical engineering from the University of Birmingham (UoB), UK, in 2025. Her research interests include reinforcement learning, multiagent systems, and their application in automotive systems.

Zhenhai Gao is a Professor with State Key Laboratory of Automotive Chassis Integration and Bionics at Jilin University, China. He received the Ph.D. degree in automotive engineering from Jilin University, 2000. He was a Postdoctoral Fellow at Xi'an Jiaotong University and a Foreign Researcher at the University of Tokyo. He has more than 20 years of research experience in a broad range of automotive engineering. In recent years, his research has been focused on connected and automated vehicles, driver behavior analysis, advanced driver assistance systems, and human-machine interfaces. More than 100 papers on his achievements have been published.