Multisource Data Fusion and Algorithm Comparison in Understanding Driving Intent

Yuqi Luo

Southampton International College (Wrexham College), Dalian Polytechnic University, Dalian, 116033, China
S23010715@mail.glyndwr.ac.uk

Abstract. This study proposes an integrated framework for detecting a vehicle's driving intention using multi-source data fusion, thereby addressing the limitations of relying on single data sources in complex traffic scenarios. Environmental perception data (other cars, pedestrians, traffic lights) are processed via YOLOv8 and BEV to extract fine-grained spatial features; vehicle motion data (speed, acceleration, steering angle) are denoised with Kalman filters to eliminate sensor drift and road unevenness interference; driver behavior insights from eye-tracking (gaze points, blink frequency) and HMI (button operations, voice commands) derive regularized features reflecting attention and intentions. The framework evaluates three algorithms: rule-driven methods offer interpretability but low accuracy, statistical learning boosts precision but depends on manual features, and deep learning achieves high accuracy yet faces latency and black box issues. Single-modality approaches lack adaptability, while multi-modal fusion reaches high accuracy but raises costs. To balance trade-offs, it suggests cross-modal lightweight integration (reducing overhead) and federated learning (protecting privacy) for algorithm selection and autonomous driving deployment, guiding practical use.

Keywords: Driving Intention, Multi-Source Data Fusion, Multi-Modal Fusion, Federated Learning, Autonomous Driving.

1 INTRODUCTION

Autonomous vehicles have become the core direction for automotive industry upgrades, with users' demands for travel safety and traffic efficiency continuously increasing. Traditional vehicle perception technologies can capture physical states, such as position and speed, but are unable to interpret lane changes, yielding deeper driving intentions, leading to decision delays or misjudgments in complex traffic interactions. This significantly hinders the commercialization of L3 and higher-level autonomous driving. Vehicle traveling intention detection, through detecting the proactive tendency of vehicles and surrounding vehicles within a 3-5 second preview period, can improve emergency response efficiency during traffic conflicts at intersections, playing a significant role in enhancing traffic system safety and efficiency.

S. Zhang (ed.), *Proceedings of the 2025 International Conference on Electronics, Electrical and Grid Technology (ICEEGT 2025)*, Advances in Engineering Research 292,
https://doi.org/10.2991/978-94-6463-986-5_58

From international research progress, vehicle traveling intention detection technology has exhibited distinct generational features. Between 2010-2015, it was dominated by rule-driven stages, with typical representatives such as the Advanced Driver-Assistance Systems (ADAS) from Bosch. The core of this stage was the "if-else" decision chain based on driving experience; through rules like steering state - steering wheel angle - lateral acceleration, intentions could be judged. Although the logic was transparent and hardware costs were low, its accuracy remained modest and the misjudgment rate notably high in complex scenarios. Between 2016-2020, it entered the statistical learning stage, with models like random forest and GBDT becoming mainstream. Through integrating multi-source features, the Baidu Apollo team improved accuracy to 80%-85%, though it still relied on manual feature engineering and had insufficient robustness in extreme environments [1]. Since 2021, deep learning has advanced significantly, with LSTM and transformer models achieving performance breakthroughs. Waymo's spatiotemporal fusion model and NVIDIA BEVFormer achieved over 90% accuracy through end-to-end learning. However, these models face high computational demands and black box decision-making security risks [2]. Overall, although technology has evolved, the system overview of different paths, applicability boundaries, performance limitations, and optimization directions remains unclear, constraining the scientific selection of techniques and targeted research breakthroughs.

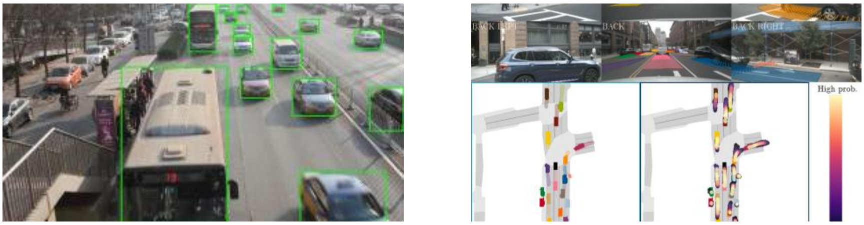
This study aims to construct a data–algorithm–scene correspondence framework for vehicle traveling intention detection, clarifying technical applicability boundaries and optimization paths. The research involves preprocessing key data categories such as environment perception, vehicle motion, and driver behavior, along with their matching logic with algorithms; analysis of the principles, cases, and performance of rule-driven, probabilistic models, machine learning, and deep learning methods to establish a multi-dimensional comparative framework for quantitative analysis. This paper follows the order of data extraction, method parsing, current trends, and conclusions, providing reference guidance for the scientific selection of intention detection techniques and promoting the commercialization of higher-level autonomous driving.

2 Data Preprocessing and Algorithm Matching

Vehicle traveling intention detection precision relies on multi-source data parsing, which requires compliance with a data characteristics-preprocessing techniques-algorithm capabilities adaptation logic. Environmental perception, vehicle motion, and driver behavior key data classes exhibit feature attribute discrepancies that necessitate the development of tailored processing and modeling schemes.

2.1 Environmental Perception Data: Spatial Modeling

Environmental perception data are used to characterize the surrounding scene of the vehicle, encompassing dynamic targets, static infrastructure, and scene structure. These core characteristics are characterized by multisource heterogeneity and strong spatial associations. Preprocessing during this stage requires addressing target parsing and spatial fusion challenges. As shown in Fig. 1, two types of environmental perception data are presented. Fig. 1(a) illustrates an example of YOLOv8 object detection results under a single sensor, whereas Fig. 1(b) demonstrates projected instance segmentations and mean future trajectories of dynamic agents. The bottom row is the future instance prediction in bird's-eye view within a 100m × 100m capture size around the ego-vehicle. The former depicts single-sensor data representation, whereas the latter illustrates the process of projecting multi-angle sensor data onto a bird's-eye coordinate system to achieve unified perspectives and spatial feature integration, making the processed data suitable for algorithms such as those used in NVIDIA BEVFormer, which fuse spatial features from camera images and LiDAR point clouds to perform intention detection and dynamic interaction judgment in complex scenarios. This multimodal fusion modeling approach provides precise spatial information support for advanced driver-assistance systems [3] [4].



(a) (b)

Fig. 1. Target Detection and BEV Spatial Fusion Diagrams: (a) Example of Target Detection using YOLOv8 ; (b) Multimodal future predictions by BEV network [3] [4].

2.2 Vehicle Motion Data: Temporal Modeling

The processed data are suitable for two types of algorithms: the Gradient Boosted Decision Tree (GBDT), which excels in detecting nonlinear relationships and short-term motion trends, and the LSTM, which shows outstanding performance in long-term previsions in complex scenarios, such as exits on highways, owing to its gated mechanism to remember long-term dependencies.

2.3 Driver Behavior Data: Intent Modeling

The Processed data is suitable for two types of algorithms: Decision Trees implement transparent and traceable reasoning through hierarchical rules, applicable to scenarios such as responsibility attribution requiring explicit explanations; Random Forests integrate multiple trees to fuse multi-dimensional features, reducing overfitting risks and enhancing the accuracy of complex intent detection.

3 Intent Detection Methods and Performance Analysis

3.1 Based on Rules

Based on rules is an early technique in vehicle intent detection. The core of this method involves transforming driving knowledge into a chain of 'if-then' rules and using feature threshold matching to determine intent. A typical example is BMW's early ADAS system for detecting lane changes, which employs a manually constructed decision tree model with inputs such as turning indicator state, steering wheel angle, and lateral acceleration to build a three-level decision logic.

The text provides specific performance metrics: an accuracy rate of 68% on clear straight roads, a single-frame delay of less than 10ms, and the ability to run on a low-end MCU (with computing power under 1TOPS), making it suitable for early entry-level cars with driving assistance features [5]. It highlights the advantages of this method: high interpretability and cost-effectiveness. The decision process is fully traced back through the rule tree, allowing engineers to identify logical flaws from logs, adhering to ISO 26262 traceability requirements, and a short model debugging cycle of just two weeks, relying on a single camera and CAN bus with costs under 500 yuan [6].

However, it also has significant limitations: scenario coverage is highly dependent on the completeness of manually defined rules, leading to high misclassification rates for unforeseen extreme scenarios; feature association modeling is rigid, using "all or nothing" threshold judgments that fail to handle fuzzy scenarios effectively; and there is a lack of adaptive learning capabilities, making it difficult to adapt to differences in driver behavior.

Finally, the text discusses optimization strategies focusing on improving dynamic adaptability: introducing scenario-triggered weight adjustment (e.g., increasing the weight of lateral acceleration to 60% in high-speed scenarios and turning indicator state to 70% in urban areas), adding redundant validation using millimeter wave radar neighbor distance features that activate alternative rule chains when camera signal reliability is below 60%, and establishing a rules feedback loop through real-world data to identify frequently misclassified scenarios, automatically generating rule supplementation recommendations to reduce maintenance costs.

3.2 Probabilistic Methods

This study introduces a probabilistic modeling approach based on quantifying the association probability between observed data - hidden intent to address uncertainty issues in intent evolution. The typical application is the use of Hidden Markov Models (HMM) for target tracking, particularly in intention-aware lane-changing detection within ADAS systems. Notably, an improved GMM-HMM model from Chongqing University has been implemented as a pilot project in the 2021 bottled car's ADAS system by Dongfeng [7].

The model leverages the lateral offset and side velocity as observed features to decode the "left turn - maintain - right turn" state sequences using the Viterbi algorithm. A detection accuracy of 95.6% is achieved within one second prior to lane-changing events. The core value of this approach lies in its ability to quantify uncertainty while capturing temporal dependencies: the outputs include confidence intervals, which exceed ±15%, thereby triggering high-risk warnings; and by modeling state transition probabilities through time dependency relationships, such as "flashlight - steering" action associations, the robustness under observed scenarios surpasses rule-based systems [8]. However, certain limitations exist: probabilistic graph inference is computationally complex, with exponential growth in computational burden as the number of states increases; delay accumulation rises to 50ms per frame; and reliance on prior distribution assumptions may lead to performance degradation of up to 15-20% when there exists significant discrepancy between real-world scenarios and training data [9]. Additionally, multi-modal feature fusion is constrained by traditional methods that primarily process image or physiological signal data in an isolated manner.

To address these challenges, the optimization strategy focuses on enhancing adaptability and efficiency: integration of Kalman filtering to smooth observation noise; dynamic correction of state transition probabilities through a predict-update mechanism to improve the accuracy of confidence intervals under extreme scenarios; construction of lightweight incremental learning modules capable of real-time updating based on new data within 100-hour intervals, without requiring full-scale retraining; extension to Dynamic Bayesian Networks (DBNs) by incorporating directed graph structures for multi-modal feature fusion and reducing computational complexity through variable conditional independence relationships, leading to a 30% improvement in computational efficiency under multi-feature scenarios.

3.3 Methods Based on Machine Learning

This section focuses on machine learning-based methods for detecting change-related intentions in Advanced Driver-Assistance Systems (ADAS), particularly within structured scenarios such as lane-changing behaviors. The discussion highlights a representative approach from a Tier1 supplier that employs XGBoost for production implementation.

The XGBoost solution leverages CAN bus data as the core input and manually designs seven key feature types, extracting statistical information through sliding windows to generate feature vectors. A 5-fold cross-validation training process is implemented, achieving an accuracy rate of 92% in highway scenarios, which has already supported the autonomous decision-making capabilities of multiple vehicles with intelligent cruise control and lane-keeping assist systems [10].

The advantages of this method include effective adaptation to structured data: tailored feature engineering enables the capture of fine-grained driving behaviors before a change occurs; XGBoost's gradient boosting and regularization strategies ensure model stability across 100,000 samples, maintaining an accuracy difference between the training and testing sets below 3%, thus avoiding overfitting; and feature importance ranking provides traceable decision-making logic, complementing the black box limitations of deep learning.

However, several limitations are evident: reliance on manually designed features constrains performance based on feature quality, leading to significant drops in accuracy (down to 75%) if critical information is omitted; re-engineering new features requires extensive workflows and iterative cycles exceeding two weeks; and limited capability in processing unstructured data results in information loss rates of 15%-20%, hindering adaptability to complex scenarios [10].

To address these challenges, the optimization strategy focuses on targeted improvements: introducing SHAP values to automatically select high-contributing features, reducing redundancy, and lowering human feature design costs from the outset; implementing an incremental learning framework by updating only the final three layers of tree structures to achieve model retraining within a single hour, significantly reducing scenario adaptation cycles; and most critically, advancing multi-modal fusion capabilities by integrating environmental perception data with vehicle motion data via feature concatenation and incorporating attention mechanisms during training to dynamically allocate modal weights. This targeted optimization directly enhances the accuracy of complex scenarios by 8%-10%, providing more comprehensive support for technical applications [10].

3.4 Methods Based on deep Learning

This section explores advanced driver-assistance systems (ADAS) that utilize deep learning methods to detect change-related intentions, particularly in lane-changing behaviors within structured scenarios, such as highway driving situations. The discussion highlights a representative implementation approach from a Tier1 supplier that employs the NVIDIA BEVFormer, an exemplar application for high-speed exit detection [11]. NVIDIA BEVFormer integrates camera images and LiDAR point clouds to perform end-to-end learning, extracting complementary three-dimensional information to optimize small-object recognition while reducing lane-changing errors through unified coordinate systems. This multi-modal feature extraction also enhances neighbor vehicle feature extraction, thereby improving the accuracy of continuous lane changes prediction.

The core advantages of such deep learning-based methods lie in their end-to-end characteristics, which provide efficiency improvements and enhanced generalization capabilities compared with traditional machine learning approaches. Traditional machine learning relies on manually designed features, requiring domain experts to invest significant effort to identify key signals [12]. In contrast, end-to-end models automate feature extraction and decision modeling from input data to output results, thereby significantly reducing the costs of human intervention. For complex scenarios, such as dynamic environments with changing obstacles or non-standard markings, end-to-end methods excel by naturally integrating multi-source information. This approach also benefits from the introduction of the Transformer architecture, which leverages self-attention mechanisms to simultaneously capture temporal and spatial relationships between data points, unlike LSTM networks, which are constrained by their chain-like structure in modeling long-range dependencies. Transformers not only enhance the stability of long-term sequence prediction but also overcome the limitations associated with structured data, enabling the direct processing of non-structured information, such as images and point clouds, while covering more environmental details.

3.5 Method Comparison and Synergy Insights

In the lane-change intention detection of Advanced Driver-Assistance Systems (ADAS), there are four core methods: based on rules, probabilistic methods, machine learning, and deep learning. Through the analysis in Table 1, this paper has identified significant differences among these methods in terms of core principles, typical accuracy rates, advantageous application scenarios, and core limitations. Furthermore, this paper has clarified the technical boundaries of each method by means of comparison.

Table 1. Comparison of Feature, Performance, and Optimization for Lane-Change Intent Detection Methods. (Data from: this study)

Method Type	Core Principle	Typical Accuracy	Advantageous Scenarios	Core Limitations
Rule-based Methods	Manual "if-then" rule matching	65%–70%	Simple scenarios, low compute	Limited scenario coverage, no autonomous learning
Probabilistic Models	Quantify uncertainty via state transition probabilities	75%–80%	Partial occlusion, mild low-light interference	High computational complexity, reliance on handcrafted features
Machine Learning	Handcrafted features + supervised learning for nonlinear correlation mining	85%–92%	Structured vehicle data scenarios	Limited scenario coverage, no self-learning ability
Deep Learning	End-to-end automatic multimodal correlation extraction	90%–95%	Complex multimodal interaction scenarios	High computational demand, black-box nature, reliance on large annotated datasets

Regarding the common limitations in lane-changing detection, optimization can be conducted from four aspects: First, the multimodal data fusion mechanism can be strengthened by organically integrating multisource information, such as vehicle motion and environmental perception. This reduces the risk of poor generalization caused by reliance on single-source data and lowers the cost of manual feature engineering through the use of automatic feature selection technology. Second, the implementation of interpretability technologies should be promoted. By combining feature importance ranking (e.g., SHAP values in machine learning) with visualization tools (e.g., Grad-CAM heatmaps in deep learning), the traceability of decision logic is achieved, breaking the trust barriers of "black-box" models. Third, the model efficiency and data utilization should be optimized. Adopt methods such as incremental learning and knowledge distillation to shorten the model iteration cycle and reduce computational overhead. Meanwhile, leverage federated learning and self-supervised pre-training to decrease annotation costs and alleviate data dependency pressure. Fourth, establish a dynamic scenario adaptation mechanism. Dynamically adjust feature weights and sensor configurations for different scenarios, such as highways, urban areas, and extreme weather, thereby improving robustness and adaptability in complex environments.

4 CONCLUSION

This study constructs a data-algorithm-scenario corresponding framework, clarifying the boundaries and optimization paths of various technical approaches. It provides theoretical support and system references for technology selection in intelligent driving intent detection, serving as a core value proposition for the research. Based on this foundation, it systematically analyzes the generational characteristics of rule-driven methods, probabilistic models, machine learning, and deep learning techniques. It traces the performance differences and applicable scenarios of these technologies from the high interpretability but low accuracy of rule-based methods to the 90%+ accuracy and computational intensity limitations of deep learning, offering a clear roadmap for technological evolution. By elucidating the preprocessing logic for environmental perception, vehicle motion, and driver behavior data, it lays the groundwork for the technical implementation. It also validates that multimodal fusion enhances robustness to 93%, while highlighting future optimization directions in lightweight integration, enhanced explainability, and reduced reliance on data. These insights guide the path toward technology deployment. Ultimately, the research directly supports the commercialization of Level-3 and higher-level autonomous driving systems, demonstrating its significant impact on industrial practice.

References

1. Bhang, W., Li, H., Li, Y., Chou, J., Xiang, H.: Semi-Supervised Hierarchical Recurrent Graph Neural Network for City-Wide Parking Availability Prediction. In: AAAI Conference Proceedings, pp. 12345-12352. AAAI Press, Menlo Park, CA, USA (2020).
2. Li, Z., Wang, W., Li, H.: BEVFormer: Learning Bird's-Eye-View Representation From LiDAR-Camera Via Spatiotemporal Transformers. IEEE Transactions on Pattern Analysis and Machine Intelligence 47(3), 2020-2036 (2025).
3. Liu, X. J., Wang, Y. P., Yu, D. X., & Yuan, Z. M.: YOLOv8-FDD: A Real-Time Vehicle Detection Method Based on Improved YOLOv8. IEEE Access 12, 136280-136296 (2024)
4. Hu, A., Murez, Z., et al.: FIERY: future instance prediction in bird's-eye view from surround monocular cameras. arXiv. arXiv 12, 136280-136296 (2024).
5. Basma, K., Antoine, T., et al.: Comparing Rule-Based and Data-Based Approaches for Lane-Change Prediction. In: Antão, P. (eds.) Proceedings of the 31st European Safety and

Reliability Conference (ESREL 2021), ESREL Conference Series on Safety, Reliability and Risk Analysis, vol. 5, pp. 1-8. CRC Press, Boca Raton, FL, USA (2021).
6. He, S., Zeng, J., Zhang, B.: Rule-Based Safety-Critical Control Design using Control Barrier Functions with Application to Autonomous Lane Change. In: IEEE Control Systems Society Conference Proceedings, pp.178–185. IEEE Press, Piscataway, NJ, USA (2021).
7. Yang, C. H., Zheng, L., Zuo, Y. F.: Highway lane-changing behavior: a data-driven analysis of driver intentions. Journal of Chongqing University 47(11), 37-50 (2024).
8. Zhang, Y., Zou, Y., Selpi, Zhang, Y., Wu, L.: Spatiotemporal Interaction Pattern Recognition and Risk Evolution Analysis During Lane Changes. IEEE Transactions on Intelligent Transportation Systems 24(6), 1-11 (2023).
9. He, G., Li, X., Lv, Y., Gao, B., Chen, H.: Probabilistic intention prediction and trajectory generation based on dynamic bayesian networks. In: IEEE Chinese Automation Congress Proceedings Series, pp. 2646-2651. IEEE Press, Piscataway, NJ, USA (2019).
10. Zhang, Y., Shi, X., Zhang, S., Abraham, A.: A XGBoost-Based Lane Change Prediction on Time Series Data Using Feature Engineering for Autopilot Vehicles. IEEE Transactions on Intelligent Transportation Systems 23(10), 19187-19200 (2022).
11. Luo, X., Huang, Y., Cui, J., Zheng, K.: Deep learning-based lane detection for intelligent driving: A comprehensive survey of methods, datasets, challenges and outlooks. Neurocomputing 650, 130795 (2025).
12. Maity, M.; Banerjee, S.; Chaudhuri, S. S.:, In: Proceedings of the 5th International Conference on Computing Methodologies and Communication (ICCMC), pp. 1442-1447. IEEE Press, Piscataway, NJ, USA (2021).