applied sciences

MDPI

Article

Assessment Scheme for Scenario Allocation in Automated Driving Based on a Hybrid Genetic–Fuzzy Framework

Botian Mei [1,2], Xiaojun Zhang [1], Hang Sun [2], Lin Zhang [3] and Yiding Hua [2,*]

[1] School of Mechanical Engineering, Hebei University of Technology, Tianjin 300401, China; xjzhang@hebut.edu.cn (X.Z.)
[2] China Automotive Technology and Research Center Co., Ltd., Tianjin 300300, China; sunhang@catarc.ac.cn
[3] School of Automotive Studies, Tongji University, Shanghai 201804, China; tongjizl@tongji.edu.cn
* Correspondence: huayiding@catarc.ac.cn; Tel.: +86-18622766087

Abstract

To address the structural differences between closed-track and open-road testing in terms of scenario coverage, risk controllability, and validation consistency, this study proposes a scenario-driven combined testing method for automated driving systems. The proposed approach constructs a multi-dimensional scenario space based on functional decomposition and jointly quantifies scenario complexity and hazard level from the perspectives of information heterogeneity and interaction-induced risks. Based on these two-dimensional scenario attributes, a fuzzy inference mechanism is developed to dynamically allocate validation resources across different testing environments. To further improve rule-base generalization and mapping stability, an enhanced genetic algorithm integrating simulated annealing and K-means clustering is introduced to optimize the rule structures in an evolutionary manner. Experimental results demonstrate that, compared with traditional testing methods and single-mechanism optimization strategies, the proposed approach achieves a more consistent and interpretable mapping between scenarios and testing proportions in high-complexity urban traffic scenarios. While ensuring test adequacy, the testing economy is significantly improved, with an overall average improvement exceeding 20%. In addition, stable resource allocation performance is observed across multiple scenarios with different levels of complexity and risk, confirming the scalability and applicability of the proposed method for multi-scenario automated driving system testing.

Keywords: scenario-based validation; fuzzy inference; automated driving systems; combined testing; improved genetic algorithm

Check for updates

Academic Editor: Christos Bouras

Received: 11 December 2025
Revised: 30 December 2025
Accepted: 7 January 2026
Published: 8 January 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

With the continuous evolution of automated driving technologies, safety validation has become a critical factor constraining their large-scale deployment. At present, the industry generally adopts a multi-pillar validation framework composed of simulation testing, proving ground testing, and open-road testing [1–4]. However, behavioral uncertainty in complex traffic environments, the diversity of traffic participants, and the high dynamic nature of scenarios make it difficult for any single testing modality to comprehensively cover the potential risk space. In particular, under heterogeneous traffic conditions, different testing environments exhibit significant differences in terms of risk exposure capability, cost structure, and controllability. How to achieve quantitative and interpretable scenario allocation among simulation, proving grounds, and open roads, and to establish a systematic evaluation scheme for testing resources, still lacks unified theoretical support.

In terms of testing platforms and testing environment frameworks, existing studies have mainly focused on the provision of testing capabilities and the construction of validation infrastructures, leading to the development of various testing platforms and scenario libraries [5]. These include large-scale simulation platforms for virtual environments and vehicle–road collaborative testing platforms [6–8], reconfigurable proving grounds and multi-level scenario libraries that support diverse road conditions and facility configurations [9,10], as well as layered validation frameworks integrating virtual testing with physical testing [11]. These studies have achieved significant progress in enhancing the operability of testing environments and the coverage of scenarios. However, their research emphasis has predominantly remained on the supply side of platforms and resources, without theoretically elucidating why different scenario attributes impose differentiated requirements on simulation-based testing, closed-proving-ground testing, and open-road testing, nor providing a unified quantitative evaluation scheme applicable to scenario allocation across different testing environments.

With regard to improving the detection efficiency of automated driving testing, existing studies have established a class of testing methodologies that take increasing the occurrence probability of high-risk conditions as the core objective [12–16]. Such studies mainly focus on the scenario exposure process itself [17], with optimization targets concentrated on scenario sampling and risk manifestation within a single testing environment. However, most existing research is conducted under a single testing environment framework and has not systematically addressed how scenarios with different levels of complexity and risk should be coordinated and allocated across multiple testing environments [18], nor has it provided a theoretical basis for a unified quantitative evaluation of such cross-environment allocation strategies.

Existing studies indicate that the performance of automated driving systems varies significantly across scenarios with different levels of complexity and risk [19,20]. They lack unified mathematical representations and quantitative comparison mechanisms, and have not established an interpretable mapping from scenario attributes to testing requirements.

Overall, within existing research frameworks, long-standing and pervasive challenges remain in terms of result consistency and validation credibility across different testing environments. Therefore, there is still no approach that simultaneously possesses:

(1) a unified quantitative representation of scenario complexity and risk;
(2) a mechanism to characterize the nonlinear relationships between scenario attributes and testing requirements;
(3) a systematic framework that supports the evaluation of scenario allocation between proving ground testing and open-road testing.

To address the above challenges, this paper investigates the coupled modeling of scenario characteristics and testing resource allocation in automated driving validation. The focus is placed on how scenario-level attributes can support coordinated testing decisions across heterogeneous environments, including closed proving grounds and open-road testing.

A scenario-driven evaluation framework is developed, in which scenario complexity and risk are jointly modeled to form a unified representation of testing requirements. This unified scenario space enables consistent quantitative descriptions of multi-source and multi-type traffic scenarios, providing a common basis for subsequent allocation decisions across different testing environments.

Based on this representation, a fuzzy inference mechanism is employed to establish an explicit mapping between scenario attributes and testing environment allocation ratios. To improve the stability of this mapping under varying scenario conditions, evolutionary

optimization is incorporated to refine the inference structure and parameter configuration while maintaining the semantic organization of the rule base.

The proposed framework is evaluated using representative highway, urban, and proving-ground scenarios, together with a cost-efficiency assessment model. The experimental results demonstrate the effectiveness of the framework in terms of testing coverage, resource utilization efficiency, and economic feasibility, indicating its applicability to multi-scenario automated driving validation.

The structure of the paper is illustrated in Figure 1.

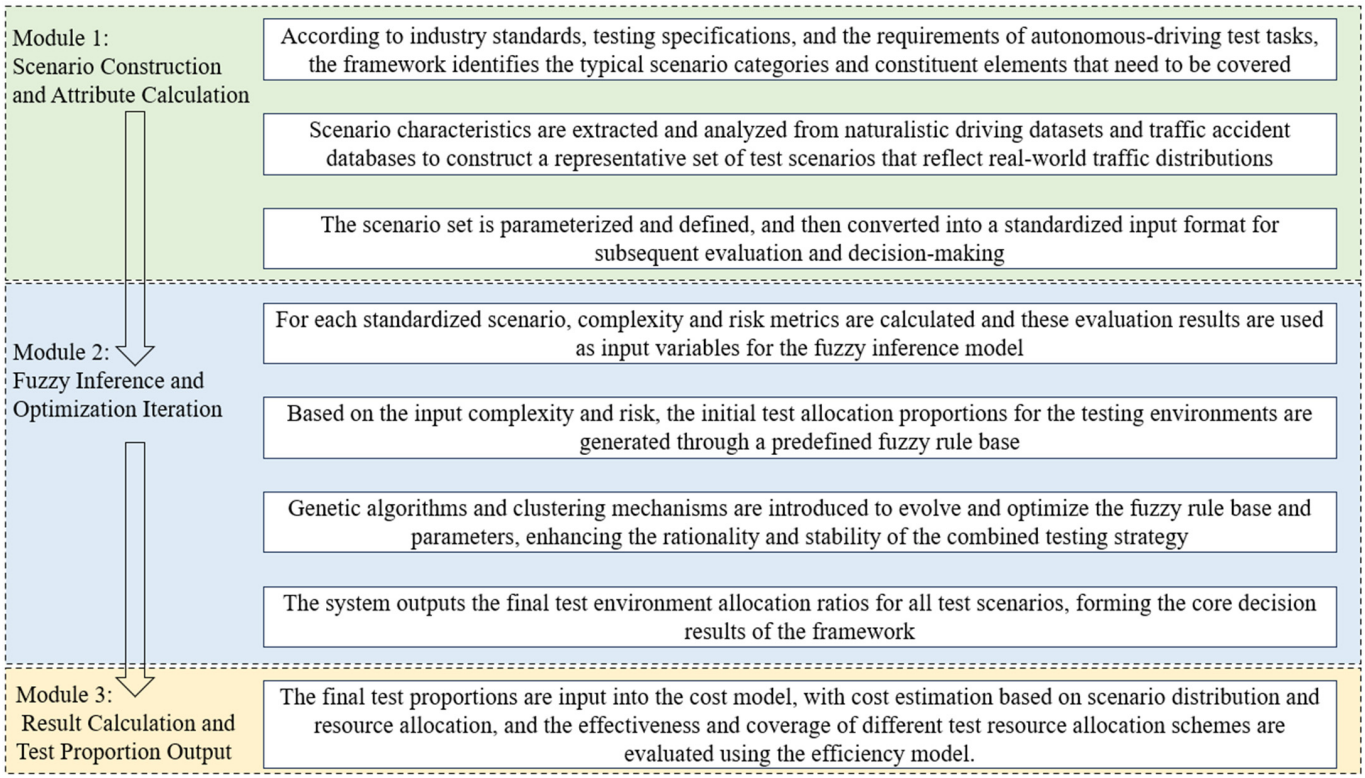


Figure 1. Research workflow diagram.

2. Materials and Methods

2.1. Scenario and Data Construction

This section constructs a testing scenario set covering three typical road environments—highways, urban roads, and proving grounds—and establishes a unified feature space for quantitative scenario characterization. The construction process is primarily data-driven. Through the joint analysis of naturalistic driving data and accident samples, key features reflecting both safety-criticality and the diversity of traffic interactions are extracted, including speed, traffic flow density, traffic participant types, and environmental disturbance conditions. In testing environment allocation for automated driving systems, scenario characterization needs to account for both aspects. Accordingly, scenario risk is introduced to represent the potential severity and uncertainty of safety outcomes, which directly constrain the admissible testing environments, as commonly considered in scenario risk assessment studies [21]. Scenario complexity captures the structural richness of traffic interactions and the combinatorial space of participant behaviors, which has been widely used to identify challenging driving scenarios based on real-world data [22]. As these two attributes are complementary and may vary independently in real-world traffic, a joint risk–complexity representation is adopted as the basis for subsequent testing environment allocation.

2.1.1. Introduction to the China-FOT and CIDAS Data Sources and Data Selection Criteria

Existing datasets used in automated driving research can generally be categorized into two types: naturalistic driving databases, which record continuous real-world driving behaviors under normal operating conditions, and accident or in-depth crash databases, which focus on safety-critical events and collision causation. Representative datasets include China-FOT and CIDAS from China, as well as the European UDRIVE database and the U.S. SHRP2 Naturalistic Driving Study (NDS). The basic differences and primary purposes of the four databases are summarized in Table 1.

Table 1. Comparison of China-FOT and CIDAS Databases.

Database Name	Source	Data Type	Coverage	Main Purpose
China-FOT	Real-road data collection from multiple provinces and cities in China	High-frequency time-series data	Driving behavior, traffic participant states, environmental information	Driving behavior modeling
CIDAS	China Road Traffic Accident Investigation System	Structured accident data	Accident types, injury severity, environmental conditions	Scenario risk modeling
UDRIVE	Multi-country naturalistic driving data collection in Europe	Naturalistic driving data with standardized annotations	Driver behavior, vehicle dynamics, contextual information	Driver behavior analysis and traffic safety studies
SHRP2 Naturalistic Driving Study (NDS)	Strategic Highway Research Program 2, United States	Long-term naturalistic driving trajectory data	Vehicle dynamics, driver inputs, roadway environment	Behavior modeling, risk assessment, and large-scale safety evaluation

UDRIVE is a multi-country European naturalistic driving dataset with standardized behavioral annotations and is mainly used for driver behavior analysis and traffic safety studies, while SHRP2 NDS is a large-scale U.S. dataset providing long-term naturalistic driving trajectories together with detailed vehicle–driver information for behavior modeling and risk assessment. China-FOT is a large-scale naturalistic driving database collected under real-world traffic conditions across multiple Chinese cities, providing high-frequency time-series data on vehicle kinematics, driver control behaviors, surrounding traffic participant states, and environmental conditions. Owing to the high traffic density, mixed traffic composition, and frequent interaction scenarios in urban China, China-FOT is particularly suitable for capturing complex multi-agent interactions and realistic behavior distributions. CIDAS is a structured national accident database that focuses on detailed accident causation analysis and contains comprehensive information on collision types, impact configurations, injury severity, road environments, and traffic participant involvement. Unlike naturalistic driving datasets, CIDAS explicitly represents safety-critical and failure scenarios, thereby enabling systematic extraction of high-risk interaction patterns and key risk factors. The two data sources are complementary in terms of temporal resolution and event types, providing a solid data foundation for constructing scenario sets that cover both routine operating conditions and high-risk situations.

In this study, a data-driven scenario analysis method [23] is adopted. After data screening, a total of 1200 representative samples are obtained. According to typical L2/L2+ functional classifications, scenarios are grouped into automated lane keeping, automated lane change, ramp entry and exit, urban lane keeping, intersection cross-

ing, low-speed driving, and automated parking. The corresponding relationships are shown in Table 2.

Table 2. Typical Intelligent Driving Function Metrics.

V2X Single-Vehicle Intelligent Driving Function	Operating Environment	Traffic Participants
Automatic lane keeping	Expressways, Urban Express Roads	Motor Vehicles
Automatic lane change		
Ramp entry and exit		
Urban lane keeping	Urban Ordinary Roads	Motor Vehicles, Non-motor Vehicles, Pedestrians
Urban lane change		
Intersection passing		
Low-speed driving	Parking Lots, Residential Areas, and Similar Enclosed Roads	Motor Vehicles, Non-motor Vehicles, Pedestrians
Automatic parking		

To ensure the representativeness and completeness of the samples, the following rules are applied during data screening:

(1) Samples must contain complete key fields, including trajectories, speed, time, weather conditions, and other essential information;
(2) Accident samples must have clearly defined collision causality and belong to vehicle–vehicle or vehicle–pedestrian interaction types;
(3) Samples with missing fields, unclear labels, or irreproducible scenarios are excluded.

2.1.2. Environmental Factors and Scenario Feature Space

The set of environmental factors listed in Table 3 is adopted to provide a unified description of road geometry, lane organization, traffic control facilities, obstacle distribution, and environmental conditions. Each category of factors is discretized into a finite set of values, which are used to construct the scenario graph nodes and attributes in the static complexity model.

Table 3. Testing Environment Elements.

Static Environmental Element	Static Environmental Element
Road alignment	Straight road, curved road, S-shaped curve, U-shaped curve
Road segment	One-way, two-way, four-way
Lane	Two-lane, three-lane, irregular lane
Traffic participants	Motor vehicles, non-motor vehicles, pedestrians
Movement direction	Same direction, opposite direction, crossing, perpendicular
Weather & time	Clear, rain, snow, daytime, nighttime, dusk
Traffic lights	Motor vehicle traffic lights, non-motor vehicle traffic lights
Obstacles	Parked vehicles, shrubs, traffic guardrails

2.1.3. Selection of Representative Parameterized Scenarios

To cover representative regions of the complexity–risk space while controlling the total number of scenarios, this study constructs a testing matrix based on the complete scenario set by selecting four key parameter dimensions: vehicle speed, target object speed, weather conditions, and traffic flow density. Through multidimensional parameter combinations,

eight representative parameterized scenarios (indexed as 1, 21, 31, 33, 34, 36, 39, and 47) are selected from 50 candidate scenarios, forming a representative subset that includes highway straight-road conditions, urban lane-keeping and lane-changing conditions, multi-agent interaction conditions at intersections, and low-speed proving-ground conditions.

Based on the environmental factors listed in Table 3 and in conjunction with functional classifications, an automated driving testing scenario set is constructed, with partial examples shown in Table 4. Scenarios are grouped by functions such as highway lane keeping, low-speed automated driving, urban lane keeping, and intersection crossing. By combining different environmental factors and participant configurations, a progressive sequence of scenarios ranging from low complexity to high complexity is formed. The scenarios indexed as 1, 21, 31, 33, 34, 36, 39, and 47 in Table 4 correspond to typical conditions such as braking of a preceding vehicle on a straight highway segment, low-speed disturbances in a proving ground, urban lane merging, and multi-agent convergence at intersections. These scenarios constitute the representative subset selected for subsequent analysis

Table 4. Autonomous Driving Test Scenario Set (Partial).

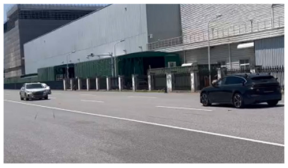
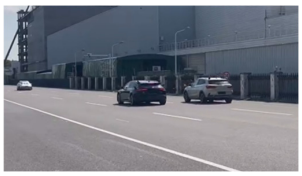
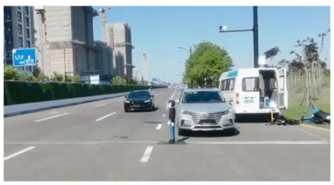
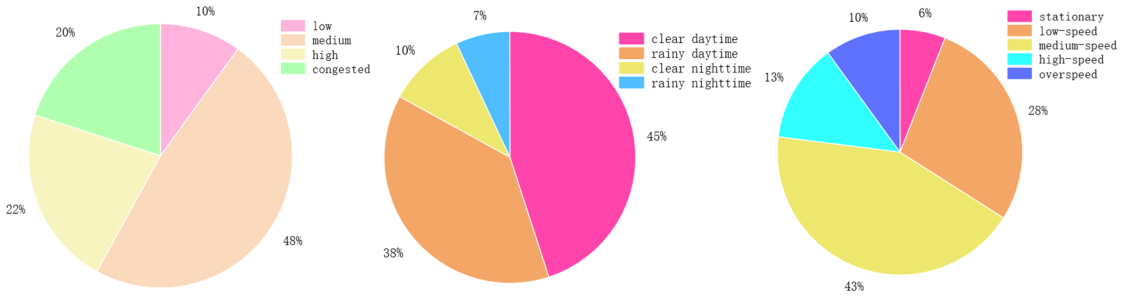
Category	ID	Name	Test Scenario Diagram
Highway lane keeping	1	Lead–lead vehicle brakes, lead vehicle cuts out	
Low-speed autonomous driving	21	Encountering non-motorized vehicles on yard roads	
. . .	. . .	. . .	. . .
Urban Lane Keeping	31	Vehicle present in adjacent lane, vehicle from the other side lane cuts in	
	33	Rear vehicle from adjacent opposing lane drives against traffic and cuts in	
	34	Non-motorized vehicle crosses ahead from adjacent lane while ego vehicle goes straight	
	36	Lead vehicle in ego lane drives against traffic, vehicle present in adjacent same-direction lane	

Table 4. *Cont.*

Category	ID	Name	Test Scenario Diagram
...	...	...	...
Intersection passing	39	Ego vehicle goes straight, pedestrian crosses from roadside obstacle	
	47	Oncoming-side occlusion with cross-traffic approaching	

Based on the statistical analysis of China-FOT naturalistic driving data and CIDAS accident records, this study characterizes the distribution features of urban road scenarios from three key environmental dimensions: vehicle speed, weather conditions, and traffic flow density. These characteristics are used to constrain parameter ranges and to guide the construction of the testing matrix. The statistical results, as illustrated in Figure 2, indicate that urban vehicle speeds are mainly concentrated in low- and medium-speed ranges, while non-ideal weather conditions and high-density or congested traffic flows account for a relatively large proportion. Complex operating conditions are often generated by the superposition of multiple factors.



(**a**) Speed Distribution (**b**) Weather Distribution (**c**) Traffic Flow Distribution

Figure 2. Distributions of key parameters in urban road scenarios based on the China-FOT and CIDAS datasets: (**a**) vehicle speed, (**b**) weather conditions, and (**c**) traffic flow density.

Based on these findings, the testing scheme adopts a non-uniform speed sampling strategy dominated by low and medium speeds, and constructs a stratified test set according to levels of environmental degradation and traffic flow density. This ensures that the selected representative scenarios more closely reflect the statistical characteristics of real-world urban traffic scenarios in terms of complexity and risk.

According to the statistical results of the databases, vehicle speeds in urban road scenarios are mainly distributed across three levels: low, medium, and high, while traffic flow density is primarily distributed across three levels: low, moderate, and high. Therefore, subsequent scenario construction adopts values based on these discrete levels to ensure consistency with real-world distributions.

Scenario design follows a progressive principle of "baseline–variation–extreme." Baseline scenarios are used to represent typical operating conditions at a moderate level of complexity. Variation scenarios adjust vehicle speed or traffic flow density while keeping

other conditions unchanged, in order to analyze the influence of single factors. Extreme scenarios introduce strong disturbance conditions, such as high-speed driving in rainy nighttime environments or congested traffic, to cover high-risk boundary cases. During the construction of the scenario matrix, factors such as the relative speed between the ego vehicle and target objects, as well as lighting conditions, are also considered. Several paired scenarios are designed to analyze the independent effects of relative speed and environmental factors on system performance, as well as their interaction effects. The detailed parameter configurations of the representative scenarios are provided in Table 5.

Table 5. Test Scenario Parameters.

ID	Ego Vehicle Speed (km/h)	Target Object Speed (km/h)	Weather Condition	Traffic Flow Density
1	80	70	Rain + night	Low
21	20	10	Clear + night	High
31	40	20	Clear + daytime	Moderate
33	40	30	Clear + daytime	Low
34	30	20	Clear + daytime	Low
36	30	10	Rain + night	Moderate
39	30	10	Clear + night	Moderate
47	30	30	Clear + daytime	Moderate

2.2. Scenario Evaluation Model

This section first constructs a scenario complexity evaluation model, in which road testing scenario complexity is decomposed into two dimensions: static complexity and dynamic complexity [24]. These dimensions, respectively, characterize the difficulty features of environmental structure and traffic interaction processes. To quantify complexity, information entropy theory is introduced based on spatiotemporal interaction relationships among vehicles, and discrete information sources within the scenario are modeled. The basic scenario complexity is then obtained through the combination of static entropy and dynamic entropy. Subsequently, on this basis, a risk evaluation model integrating dynamics- and risk-based indicators is developed to reflect the intensity of potential collision risk in the scenario.

2.2.1. Static Complexity

To reduce the influence of subjective experience on scenario complexity evaluation, information entropy theory is introduced to quantify static scenarios. Static complexity adopts a two-layer entropy-weighted structure: the first layer uses information entropy to characterize the diversity and non-uniformity of scenario elements, while the second layer reflects the relative importance of different elements to the driving task through weighting. Under a graph-based representation, the basic model of static scenario complexity is defined as:

$$\theta = -\sum_{i=1}^{n} p_i log_2 p_i \quad (1)$$

where θ denotes the static scenario complexity coefficient, n represents the total number of scenario element groups, and p_i indicates the proportion of nodes of a specific type in the graph structure relative to the total number of nodes. Unlike classical Shannon information entropy, physical elements such as roads, buildings, and traffic signs are mapped to nodes in a graph structure, and connections are established based on their

topological relationships. In this way, the complexity of static scenarios is quantified in the form of structural entropy. Considering that different elements have varying degrees of influence on driving decision-making, an entropy-weighted corrected static complexity is introduced:

$$C_1 = \frac{\sum_{i=1}^{h}(\beta_i \sum Y_i)}{\theta} \tag{2}$$

where C_1 denotes the static scenario complexity, β_i represents the weight corresponding to the i-th group of elements in the static scenario composition, Y_i denotes the sum of predefined scores of all elements within that group, and β is the base information entropy. The weighting coefficients are determined through a combination of expert scoring and empirical analysis. While preserving the objective characterization of the intrinsic complexity of the scenario provided by the entropy measure, this approach also reflects the differentiated contributions of various elements to the driving task. Consequently, the static structures of different road environments can be compared under a unified metric.

2.2.2. Dynamic Scenario Complexity

Dynamic scenario complexity is used to characterize the interaction intensity among traffic participants and the nonlinear characteristics arising from variations in geometric relationships and motion states. Based on the statistical analysis of typical interaction scenarios, a multi-factor coupled formulation is adopted to construct the dynamic complexity model. The core form of the dynamic complexity model is given in [25]:

$$C_2 = e^{f_e f_v f_d} \tag{3}$$

where f_θ, f_v, and f_d denote the complexity factors corresponding to encounter angle, relative speed, and relative distance, respectively. An exponential mapping is adopted so that complexity increases in an accelerated manner with interaction intensity, which is consistent with the variation characteristics of driver perceptual load under critical operating conditions.

The angle factor f_θ is adapted from angle-based risk assessment concepts in the maritime domain and modified for road scenarios. Typical encounter relationships—such as same-direction, opposite-direction, and perpendicular crossing—are mapped to different complexity weights. Perpendicular crossings are assigned the highest weight, followed by opposite-direction encounters, with same-direction encounters assigned the lowest weight, reflecting differences in the difficulty of the perception–decision chain under different intersection geometries.

The relative speed and distance factors f_v and f_d introduce an adaptive safety distance on the basis of the traditional Time To Collision (TTC) metric, incorporating vehicle dynamic characteristics and road adhesion conditions into the modeling framework. For different types of traffic participants, including pedestrians, non-motorized vehicles, and motor vehicles, differentiated distance thresholds are applied to reflect differences in risk sensitivity.

Static complexity and dynamic complexity jointly determine the basic complexity of a scenario:

$$C_B = C_1 C_2 \tag{4}$$

The multiplicative formulation is used to characterize the synergistic amplification effect between static environments and dynamic interactions, whereby identical dynamic behaviors result in higher overall complexity in more structurally complex environments.

The final scenario complexity C is determined jointly by the basic complexity and the traffic flow density modulation factor:

$$C = \xi C_B \tag{5}$$

where ξ denotes the traffic flow density. Under low traffic flow conditions, ξ approaches 1, while under moderate and high traffic flow conditions, it appropriately amplifies the complexity level. This layered modulation mechanism enables flexible adjustment of the overall complexity level across different traffic flow environments while maintaining the structural stability of the complexity model, thereby providing a quantitative basis for describing the risk amplification effect in high-interaction scenarios.

2.2.3. Risk Evaluation Method

This study proposes a collision risk measurement framework based on dynamic models. By integrating the Physical Crash Severity Index (PCSI) with the TTC, a more comprehensive scenario risk evaluation system is established. Analysis indicates that traditional single-indicator evaluation methods are insufficient to accurately characterize potential risks in complex traffic scenarios; therefore, a composite risk index D is introduced [26]:

$$D = \frac{PCSI}{TTC} \tag{6}$$

This ratio design reflects the coupling between two key dimensions: PCSI quantifies the physical severity of a collision once it occurs, while TTC characterizes the urgency of collision occurrence. Taking their ratio enables the risk assessment to simultaneously account for both dimensions, providing a more comprehensive risk characterization in scenarios where a single indicator cannot adequately capture both severity and urgency.

For TTC modeling, considering the prevalent acceleration and deceleration behaviors in automated driving scenarios, relative acceleration is incorporated into the prediction framework. A piecewise function is employed to distinguish between accelerating and decelerating cases, thereby avoiding the underestimation of risk caused by the traditional constant-velocity assumption under highly dynamic operating conditions. The calculation formula is given as:

$$TTC = f\left(d_{rel}, v_{rel}, a_{rel}, d_{safe}\right) \tag{7}$$

where d_{rel} denotes the relative distance, v_{rel} denotes the relative speed, α_{rel} denotes the relative acceleration, and d represents the safe stopping distance.

The potential collision severity index (PCSI) integrates key physical variables such as collision angle, relative speed, and mass ratio, and is formulated as a weighted combination:

$$PCSI = w_\theta F_\theta(\Delta\psi) + w_v F_v(\Delta v, d) + w_m F_m(m_1, m_2) \tag{8}$$

where F_θ, F_v, and F_m denote the angle factor, the relative speed–distance factor, and the mass ratio factor, respectively, and w_θ, w_v, and w_m are the corresponding weights. The angle factor characterizes differences in vehicle structural deformation and occupant injury under different impact angles, explaining why lateral collisions are generally more hazardous than same-direction rear-end collisions. The relative speed–distance factor reflects the physical principle that collision energy increases rapidly with increasing relative speed and decreasing separation distance. The mass ratio factor is derived from the principle of momentum conservation and characterizes the mechanism by which lighter vehicles are exposed to higher risk in asymmetric collisions. The weights are obtained through regression analysis on accident databases, among which the speed-related factor has the

highest weight, consistent with the physical law that kinetic energy is proportional to the square of velocity.

By integrating TTC and PCSI, the risk model simultaneously characterizes the urgency of collision occurrence and the potential severity within a unified framework, thereby providing risk indicators with clear physical meanings as inputs for the subsequent fuzzy system.

2.2.4. Normalization of Scenario Risk and Complexity

Before being used as inputs to the fuzzy inference system, the quantified scenario complexity and risk indices are normalized to a common numerical scale. For each index, the minimum and maximum values are determined from the entire constructed scenario dataset, and all scenario samples are linearly normalized to the interval [0, 1] using min–max scaling. The resulting normalized complexity and risk values are then directly used as the input variables of the fuzzy inference model for testing environment allocation.

2.3. Fuzzy Framework

In automated driving testing scenarios, issues such as sensor noise, incomplete information, and fuzzy scenario boundaries are pervasive, making it difficult for analytical models based on precise thresholds to stably characterize the nonlinear relationships between complex scenario attributes and testing requirements. Therefore, a fuzzy inference system [27] is employed to map scenario complexity C and risk D to the combined allocation ratios of proving-ground testing and open-road testing, thereby constructing an interpretable combinatorial testing strategy model.

Among various fuzzy inference structures [28], the Mamdani-type system expresses knowledge in the form of "IF–THEN" rules and is suitable for complex decision-making problems with multiple inputs and single or multiple outputs. It can directly reflect the influence of expert experience on decision-making. Considering both interpretability and implementation complexity, a Mamdani-type fuzzy inference structure is adopted in this study to construct the combinatorial testing strategy model.

2.3.1. Mamdani-Type Fuzzy Inference System

The Mamdani-type fuzzy inference system adopted in this study is illustrated in Figure 3 and mainly consists of four modules: fuzzification, rule base, fuzzy inference, and defuzzification. The system inputs are the scenario complexity C and risk D, while the output is a composite indicator representing the allocation ratios between proving-ground testing and open-road testing, which is used to guide testing resource allocation.

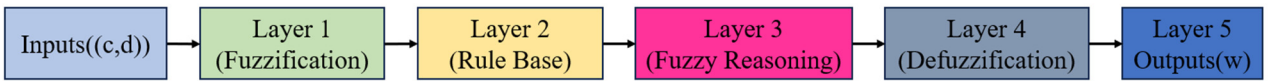


Figure 3. Mamdani-Type Fuzzy Inference System.

The fuzzification module maps the continuous input variables C and D into fuzzy sets through membership functions. Considering that variations in complexity and risk exhibit smooth transition characteristics, Gaussian membership functions are adopted to describe the input variables. This enables continuous transitions between adjacent levels at their boundaries, thereby enhancing robustness to measurement errors and parameter uncertainties.

The rule base module consists of a set of fuzzy rules in the form of IF–THEN statements. Based on combinations of complexity and risk levels, these rules provide corresponding recommendations for testing allocation ratios. The rule design integrates

expert knowledge with statistical data analysis and covers the main combination patterns within the complexity–risk space, ensuring reasonable decision outputs under typical operating conditions.

The fuzzy inference module employs the Mamdani-type minimum–maximum inference mechanism to activate the corresponding rules according to the input membership degrees and aggregates the outputs of all activated rules to obtain a fuzzy output set for the testing allocation ratio. The defuzzification module uses the centroid method to convert the fuzzy output set into a crisp numerical value, yielding the recommended allocation ratio between proving-ground and open-road testing for a given scenario. The centroid method exhibits smooth response characteristics near decision boundaries, which helps avoid abrupt changes in testing allocation ratios between adjacent scenarios.

Through the above fuzzy inference structure, the combinatorial testing strategy model realizes an interpretable mapping of "complexity–risk–testing allocation" within a unified framework, providing a foundation for subsequent rule and parameter optimization based on evolutionary algorithms.

2.3.2. Establishment of the Fuzzy-Based Combinatorial Testing Model

On the basis of the proposed Mamdani-type fuzzy inference structure, this section presents the membership functions, rule formats, and defuzzification expressions adopted in this study, and provides a formal description of the mathematical implementation of the combinatorial testing model.

(1) Inputs, Outputs, and Membership Functions

The model inputs are scenario complexity C and risk D, and the output is the proving-ground testing proportion P. The variables C, D, and P all adopt a seven-level linguistic variable set:

$$\{VL, L, LL, M, LH, H, VH\} \tag{9}$$

After discretization, a total of 7×7 scenario state combinations are formed in the (C, D) space.

As shown in Figure 4, for each input variable $x \in \{C, D\}$, the membership function of its k-th linguistic level adopts a Gaussian form:

$$u_k(x) = e^{-\frac{(x-c_k)^2}{2\sigma_k^2}} \tag{10}$$

where c_k and σ_k are the center and width parameters of the k-th linguistic level, respectively. By constraining adjacent levels to satisfy the following condition at their intersection points:

$$u_k(x_{int}) \approx u_{k+1}(x_{int}) \approx 0.5 \tag{11}$$

smooth transitions and appropriate overlap between linguistic levels are ensured, providing a tunable parameter space for subsequent evolutionary optimization.

(2) Rule Structure and Inference Mechanism

The rule base consists of a set of IF–THEN rules, and the i-th rule is expressed as:

$$R_i : IF\ C\ is\ C_i\ AND\ D\ is\ D_i\ THEN\ P\ is\ P_i \tag{12}$$

where C_i, D_i, and P_i are the corresponding linguistic levels. For a given input (C, D), the activation strength of the i-th rule is given by:

$$w_i = \min\left(u_{C_i}(\mathrm{C}), u_{D_i}(\mathrm{D})\right) \tag{13}$$

and the consequent membership function $\mu_{P_i}(p)$ is clipped accordingly:

$$u_i(p) = \min\left(w_i, u_{C_i}(\mathrm{p})\right) \tag{14}$$

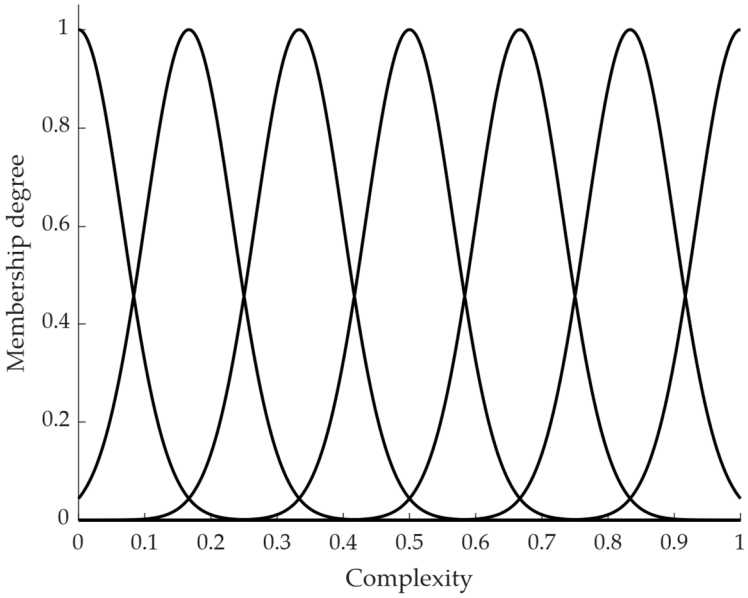


Figure 4. Gaussian Membership Functions Defining the Seven-Level Linguistic Partition for All Input and Output Variables in the Fuzzy Inference System.

All rule outputs are aggregated through a maximum operation to obtain the overall membership function of the output variable P:

$$u_{agg}(p) = \max u_i(p) \tag{15}$$

(3) Defuzzification and Output Generation

The precise value of the testing proportion is obtained through defuzzification using the centroid method:

$$P = \frac{\int_{p_{max}}^{p_{max}} p u_{agg}(p) dp}{\int_{p_{max}}^{p_{max}} u_{agg}(p) dp} \tag{16}$$

where $[p_{\min}, p_{\max}]$ denotes the domain of the output variable. To reduce computational complexity, the output domain is discretized into a finite set of sampling points p_j in the numerical implementation, and a discrete centroid formulation is adopted:

$$P = \frac{\sum_j p_j u_{agg}(p_j)}{\sum_j u_{agg}(p_j)} \tag{17}$$

Through the above definitions, scenario complexity C and risk D are mapped to a specific testing proportion P via the fuzzy rule base and membership functions, thereby forming an interpretable nonlinear mapping structure from scenario attributes to testing resource allocation.

2.4. Optimization of the Fuzzy Rule Base

In the fuzzy inference model, the rule base determines the mapping relationship between scenario complexity C, risk D, and the testing proportion P. To improve the adaptability and decision accuracy of the combinatorial testing model across different scenarios, evolutionary optimization is applied to the Mamdani rule base.

Considering that both inputs and outputs are discrete linguistic variables, the rule base is represented using a binary encoding scheme [29]. Based on the input and output linguistic levels, the rule base can be expressed as a fixed-dimension rule matrix. After binary expansion, an individual can be represented as a binary string of length L. The genetic algorithm treats this representation as the search space and iteratively updates the rule structure through selection, crossover, and mutation operators.

$$\underbrace{001}_{R_c^1}\underbrace{001}_{R_d^1}\underbrace{001}_{R_p^1}\underbrace{001}_{R_w^1}\underbrace{001}_{R_a^1}\cdots\underbrace{001}_{R_c^{49}}\underbrace{001}_{R_c^{49}}\underbrace{001}_{R_c^{49}}\underbrace{001}_{R_c^{49}}\underbrace{001}_{R_c^{49}} \tag{18}$$

The fitness function is jointly composed of the testing efficiency metric $\chi(x)$ and the resource consumption metric $E(x)$:

$$j(x) = \frac{\chi(x)}{E(x)} \tag{19}$$

The former measures testing coverage and convergence efficiency under a given rule configuration, while the latter reflects the corresponding testing time and resource consumption. Through the combined fitness design, efficiency and cost are explicitly balanced during the evolutionary process, providing a quantitative optimization objective for the combinatorial testing strategy.

To address the issues of premature convergence and insufficient population diversity that traditional genetic algorithms tend to exhibit in high-dimensional rule spaces [30], an adaptive evolutionary mechanism integrating simulated annealing and K-means clustering is introduced into the genetic framework. First, a simulated annealing operator is embedded in the genetic iteration process, where the temperature parameter T_k controls the acceptance probability of suboptimal solutions, and an exponential decay strategy is adopted for temperature updating:

$$T_{i+1} = kT_i \tag{20}$$

where T_{i+1} denotes the system temperature at the $(i+1)$-th iteration, and k is the cooling coefficient.

During the high-temperature stage, individuals with lower fitness are accepted with a certain probability to enhance the global search capability. State transitions are governed by an improved Metropolis criterion, which allows newly generated individuals with slightly lower fitness than the current individual to be accepted with a nonzero probability at higher temperatures, thereby effectively reducing the risk of being trapped in local optima.

$$P = \begin{cases} 1 & , \ J_{new} \geq J_{now} \\ e^{\frac{(Jnew - Jnow)}{T}} & , \ J_{new} < J_{now} \end{cases} \tag{21}$$

where P is the probability that the current individual is replaced by the new individual, J_{new} is the fitness of the newly generated individual, and J_{now} is the fitness of the current individual.

Second, K-means clustering is performed on the population at each generation, and the standard deviation of the distances between individuals and each cluster center is computed as a measure of population diversity. When the standard deviation is large, the population distribution is relatively dispersed and diversity is high; in this case, the crossover rate and mutation rate are appropriately increased to enlarge the search space. When the standard deviation is small, the population tends to be concentrated and close to the convergence region; accordingly, the crossover rate and mutation rate are reduced to avoid excessive disturbance to the current favorable structures and to accelerate convergence. Through

this diversity-driven adaptive adjustment mechanism, a search strategy is realized that emphasizes global exploration in the early stage and strengthens local exploitation in the later stage.

Within the improved genetic algorithm framework, both the structure and parameters of the fuzzy rule base are automatically evolved under the joint efficiency–cost optimization objective. This reduces the subjectivity introduced by manual parameter tuning and enhances the adaptability and robustness of the combinatorial testing strategy across scenarios with varying levels of complexity and risk.

2.5. Cost and Efficiency Calculation Model

To explicitly balance testing resource consumption and defect exposure capability during rule optimization, this study constructs an evaluation model for the combinatorial testing strategy from the two dimensions of cost and efficiency.

2.5.1. Testing Cost Model

The cost of proving-ground testing E_c mainly consists of site rental fees. The cost of open-road testing E_o includes three components: equipment usage cost E_{of}, personnel input cost E_{om}, and vehicle operating cost E_{ow}:

$$E_o = E_{of} + E_{om} + E_{ow} \tag{22}$$

where the equipment cost is estimated based on typical sensor rental prices reported in the literature [31], the personnel cost is calculated according to the hourly labor cost of technical staff involved in the actual testing organization, and the vehicle wear-and-tear cost is determined based on the testing intensity.

2.5.2. Scenario Exposure and Time Consumption

Unlike proving-ground testing, where operating conditions are controllable, the duration of open-road testing is strongly influenced by factors such as traffic flow, weather, and road conditions, and the occurrence of target scenarios is inherently stochastic. To estimate the time consumption required to complete testing for a specific scenario on open roads, this study introduces scenario exposure to characterize the expected number of occurrences of a given type of scenario per unit time.

For a given scenario, its exposure can be estimated based on statistical frequency:

$$p_i = f_i = \frac{n_i}{N} \tag{23}$$

where n_i is the number of occurrences of the i-th value in all statistics, N is the total number of observations, f_i is the frequency of occurrence of the i-th value, and p_i is the probability of occurrence of the i-th value.

By selecting samples with structural similarity to Scenario 47 from the IND dataset and counting their occurrences across different intersections and observation periods, the frequency distribution of Scenario 47 can be obtained. The statistical results indicate that the occurrence frequencies of most scenarios exhibit a concentrated distribution. Figure 5 presents the Kolmogorov–Smirnov test results for the occurrence frequency of Scenario 47.

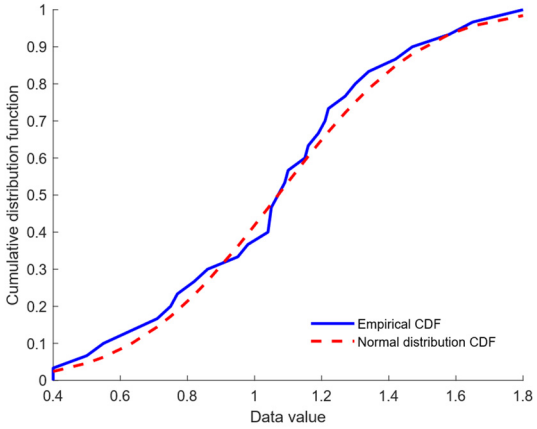


Figure 5. Kolmogorov–Smirnov Test Comparing the Empirical and Normal Cumulative Distribution Functions for the Occurrence Frequency of Scenario 47.

As shown in Figure 5, the KS test results indicate that, within this dataset, the occurrence frequency of Scenario 47 can be approximately regarded as following a normal distribution. Therefore, a normal distribution is adopted as an approximate model for scenario exposure, which smooths the influence of extreme samples on the estimation of scenario exposure time. Its probability density function can be expressed as:

$$N(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{(x-\mu)^2}{2\sigma^2}\right)} \tag{24}$$

where σ^2 and μ denote the variance and the mean, respectively.

Accordingly, the calculation formula for testing time and scenario exposure is given by:

$$Z_{0.05} = \frac{N - \mu t}{\sigma\sqrt{t}} \tag{25}$$

where $Z_{0.05}$ is the standardized Z-score corresponding to a 95% confidence level, N denotes the scenario exposure, and t represents the time required to reach the expected exposure level, satisfying:

$$t = f(p) \tag{26}$$

where $f(p)$ is a function related to the testing proportion output by the fuzzy framework.

2.5.3. Testing Efficiency Model

In modeling testing efficiency, considering that defect exposure exhibits a typical diminishing marginal return pattern as the number of tests increases, an exponential saturation model is adopted to fit the relationship between testing efficiency and the number of tests:

$$\chi(N) = \chi_{max}\left(1 - e^{-\lambda N}\right) \tag{27}$$

where $\chi(N)$ denotes the cumulative efficiency achieved after N tests, $\chi_{\max}$ is the maximum attainable testing efficiency, and λ is the efficiency constant reflecting the rate of efficiency improvement. This model captures the marginal effect characteristics of testing efficiency and can be used to evaluate the efficiency performance of different combinatorial testing strategies under given cost constraints.

Based on the above cost and efficiency models, the composite fitness function in rule base optimization can explicitly balance efficiency improvement and resource consumption, thereby providing a quantitative basis for the evolutionary optimization of combinatorial testing proportions.

3. Results

This section provides illustrative examples to facilitate a clearer understanding of how the proposed framework allocates testing requirements. Expert data are used to initialize the fuzzy rules, and the fuzzy inference process of the framework is explained in detail. Figure 6 presents the detailed algorithm employed in this study.

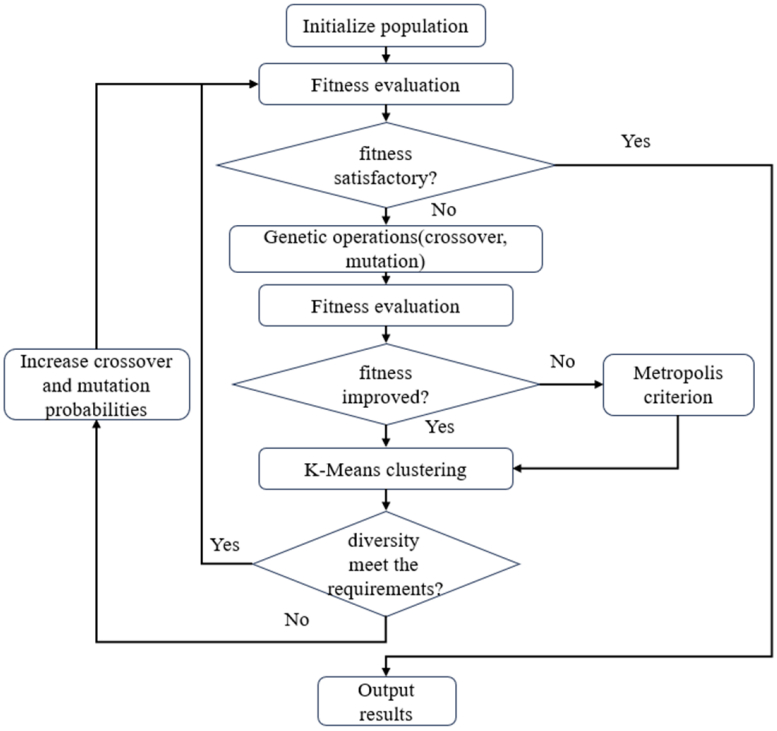
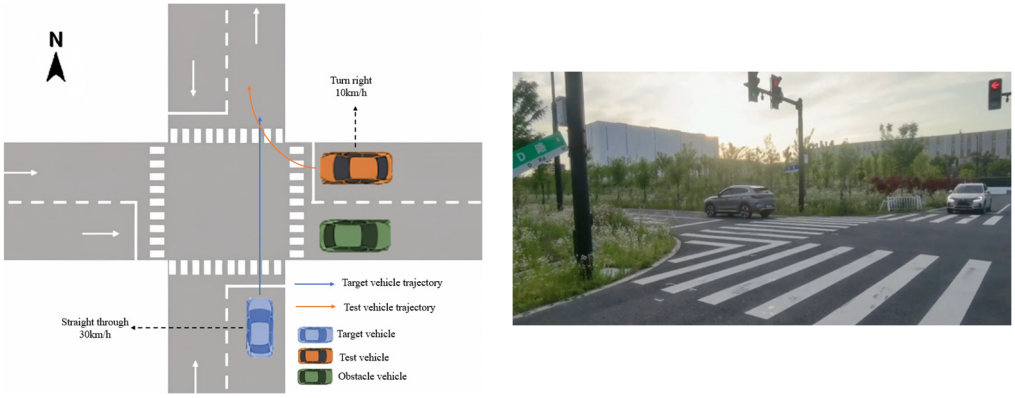


Figure 6. Flowchart of the Improved Genetic Algorithm.

3.1. Typical Scenarios and Data Sources

Eight representative parameterized scenarios—indexed as 1, 21, 31, 33, 34, 36, 39, and 47—are selected from the scenario set constructed in Section 2 of this paper. These scenarios cover three types of road environments: highways, urban roads, and proving grounds, and include both routine operating conditions and edge high-risk conditions. They are used as a unified case study set for the complexity and risk models, as well as for the combinatorial testing strategy.

Among them, Scenario 47, as illustrated in Figure 7, represents a typical high-risk urban intersection scenario. In this scenario, the ego vehicle and the target vehicle travel on perpendicular intersecting roads, with a stationary obstacle in between causing occlusion. This scenario comprehensively embodies elements such as occlusion, multi-agent interaction, and perpendicular intersection geometry, and thus has strong representativeness. Accordingly, it is emphasized in subsequent analyses of scenario exposure estimation and risk validation.



(**a**) schematic diagram (**b**) actual testing condition

Figure 7. Schematic and real-world representations of Scenario 47. (**a**) shows a schematic diagram of Scenario 47, and (**b**) presents the actual testing condition of Scenario 47.

3.2. Scenario Evaluation Metrics

In this section, taking the typical urban intersection Scenario 47 as an example, specific calculation results are presented based on the aforementioned complexity and risk models. The numerical characteristics of static complexity, dynamic complexity, and overall risk are demonstrated, and corresponding analyses are conducted in conjunction with the structural and dynamic properties of the scenario.

3.2.1. Static Scenario Complexity

The static environmental elements of Scenario 47 consist of road geometry, traffic direction structure, lane organization patterns, visible obstacles, traffic control facilities, dynamic interaction objects, relative motion patterns, and environmental disturbance factors. The hierarchical structure of these elements is illustrated in Figure 8:

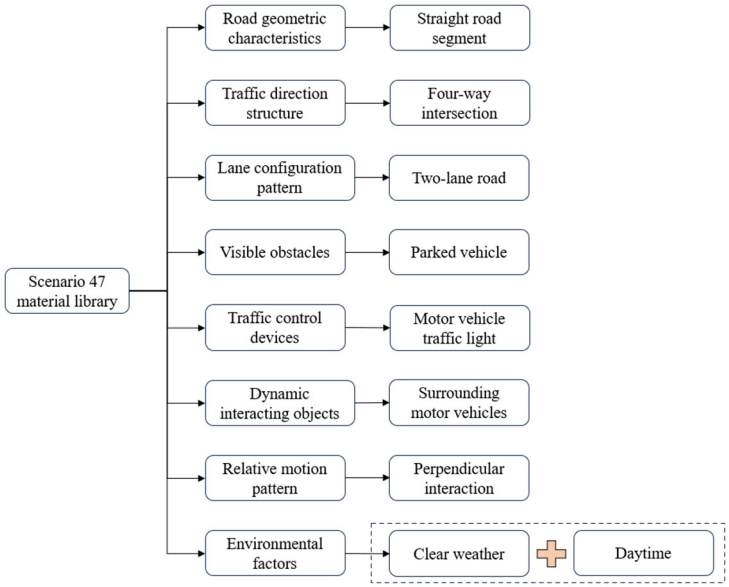


Figure 8. Static Environment Element Library.

Under the graph-based representation, the first layer distinguishes between the base material library and the scenario instance. The second layer divides road geometry features, traffic direction structures, lane organization patterns, visible obstacles, traffic control facilities, dynamic interaction objects, relative motion patterns, and environmental disturbance factors into multiple element groups, with weighting coefficients determined through a

combination of eye-tracking experiments and expert evaluation results. In the third layer, each specific element is assigned a discrete score by experts.

By substituting the element weights and element scores into the static complexity model, the static complexity index of Scenario 47 is obtained as $C_1 = 2.458$. This value is significantly higher than that of highway straight-road scenarios and simple proving-ground scenarios, corresponding to the structural characteristics shown in Figure 7, such as multi-directional traffic convergence, the presence of occluding obstacles, and the superposition of traffic control facilities. These results indicate that Scenario 47 exhibits a high level of static environmental complexity.

3.2.2. Dynamic Scenario Complexity

The experimental data of Scenario 47 are presented in Table 6:

Table 6. Experimental Data of Scenario 47.

f_θ	f_v	f_d	ξ	d_{rel} (m)	v_{rel} (m/s)	a_{rel} (m/s^2)	d (m)
0.32	0.221	0.16	1	20	30	3	1

Based on the parameters such as encounter angle, relative speed, and relative distance listed in Table 6, substitution into the dynamic complexity model yields a dynamic complexity index of $C_2 = 1.12$ for Scenario 47.

This result reflects the significant contributions of perpendicular intersection geometry, perception difficulties under occlusion conditions, and synchronous multi-agent motion to interaction complexity. By integrating static complexity and dynamic complexity, and further incorporating the modulation factor corresponding to the traffic flow density of the scenario, the overall scenario complexity of Scenario 47 is obtained as $C = 2.46$. After normalization, the standardized complexity of Scenario 47 is obtained:

$$C' = \frac{C - C_{min}}{C_{max} - C} = 0.872$$

3.2.3. Scenario Risk

According to the scenario risk model given in the equations, and based on parameters such as the relative speed, relative distance, relative acceleration, and vehicle mass between the ego vehicle and the target vehicle in Scenario 47, the risk index is obtained through calculation:

$$D' = \frac{D - D_{min}}{D_{max} - D} = 0.291$$

3.3. Fuzzy Inference

3.3.1. Rule Base

In this section, the fuzzy rules are initialized based on expert knowledge. Seven fuzzy sets are defined for each input and the output:

$$\{VL, L, LL, M, LH, H, VH\}$$

Based on the above fuzzy sets, the initial fuzzy rule base is obtained from expert knowledge, as illustrated in Figure 9:

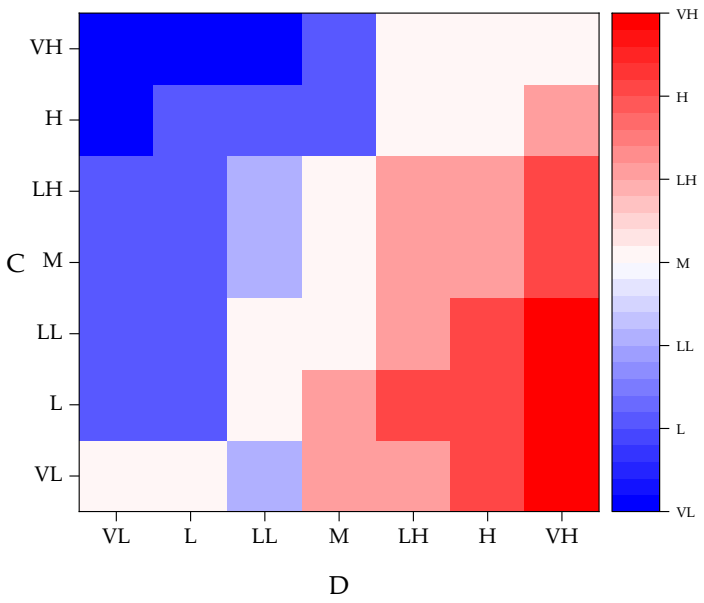


Figure 9. Fuzzy Rule Heatmap for Combined Testing of Intelligent Vehicles.

After obtaining the initial fuzzy rule base, the method proposed in this study is applied to optimize it.

3.3.2. Fuzzification

The membership degree of each input is calculated using Gaussian membership functions. For Scenario 47:

$$u_h(c) = e^{-\frac{(0.872-0.8333)^2}{2\times 0.07078^2}} = 0.7091$$

$$u_{vh}(c) = e^{-\frac{(0.872-1)^2}{2\times 0.07078^2}} = 0.1951$$

$$u_l(d) = e^{-\frac{(0.291-0.1667)^2}{2\times 0.07078^2}} = 0.2141$$

$$u_{ll}(d) = e^{-\frac{(0.291-0.3333)^2}{2\times 0.07078^2}} = 0.8365$$

3.3.3. Rule Activation Strength

The activation strength of each triggered rule is calculated using the minimum (min) operator:

$$Rule\ 1: (c\ is\ h\ and\ d\ is\ l, p\ is\ l), w = \min\ (u_h(c), u_l(d)) = 0.2141$$

$$Rule\ 2: (c\ is\ h\ and\ d\ is\ ll, p\ is\ l), w = \min(u_h(c), u_{ll}(d)) = 0.7091$$

$$Rule\ 3: (c\ is\ vh\ and\ d\ is\ l, p\ is\ vl), w = \min(u_{vh}(c), u_l(d)) = 0.1951$$

$$Rule4: (c\ is\ vh\ and\ d\ is\ ll, p\ is\ vl), w = \min(u_{vh}(c), u_{ll}(d)) = 0.1951$$

3.3.4. Defuzzification

Each rule is truncated accordingly:

$$Rule1: u_1(p) = \min(w_1, u_1(p)) = \min(0.2141, u_1(p))$$

$$Rule2: u_2(p) = \min(w_2, u_1(p)) = \min(0.7091, u_2(p))$$

$$Rule3: u_3(p) = \min(w_3, u_1(p)) = \min(0.1951, u_3(p))$$

$$Rule4: u_4(p) = \min(w_4, u_1(p)) = \min(0.1951, u_4(p))$$

For rules that share the same consequent, two rules pointing to l or two rules pointing to vl, their activation strengths can be merged first:

$$w_l = \max(w_1, w_2) = \max(0.2141, 0.7091) = 0.7091$$

$$w_{vl} = \max(w_3, w_4) = \max(0.1951, 0.1951) = 0.1951$$

Thus

$$u'_l(p) = \min\left(0.7091, u_l(p)\right)$$

$$u'_{vl}(p) = \min\left(0.1951, u_{vl}(p)\right)$$

$$u_{out}(p) = \max\left(u'_l(p), u'_{vl}(p)\right)$$

The defuzzified output is:

$$p = \frac{\int_0^1 p u_{out}(p) dp}{\int_0^1 u_{out}(p) dp} = 0.177$$

4. Comparative Analysis of Results

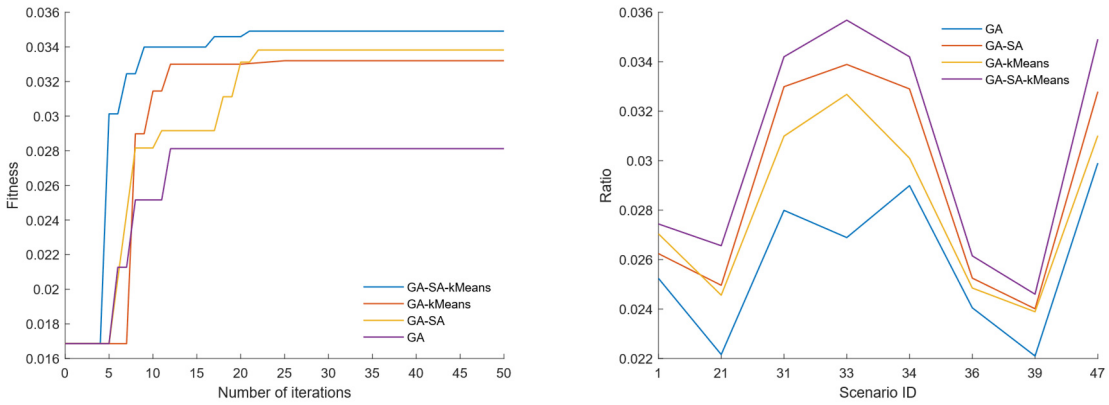
Based on the clearly defined experimental setup and evaluation metrics, this section conducts a comparative analysis of the performance of four types of genetic optimization strategies from two perspectives: fitness convergence characteristics and efficiency–cost ratio.

4.1. Ablation Study

To quantitatively evaluate the impact of different optimization mechanisms on the performance of the combinatorial testing strategy, four comparative experimental schemes based on the ablation concept are designed in this study. The specific settings of the four optimization strategies are as follows:

(1) Original genetic algorithm (GA);
(2) Genetic algorithm with simulated annealing mechanism (GA-SA);
(3) Genetic algorithm with K-means clustering-based adaptive parameter tuning (GA-K Means);
(4) Complete optimization scheme integrating simulated annealing and K-means clustering (GA-SA-K Means).

Figure 10a illustrates the overall convergence characteristics of the four strategies. The original GA exhibits slower convergence and lower terminal fitness, with evident premature convergence. Introducing the K-means mechanism (GA-K Means) significantly accelerates fitness improvement during the initial search phase but offers limited enhancement in terminal fitness. Incorporating the simulated annealing mechanism (GA-SA) primarily improves convergence quality in the mid-to-late stages. The complete scheme, GA-SA-K Means, demonstrates superior convergence performance across both initial and later phases, accelerating convergence while achieving the highest terminal fitness, reflecting the synergistic benefits of the two mechanisms in global exploration and local exploitation.



(**a**) Fitness convergence behavior (**b**) Efficiency-to-cost ratios

Figure 10. Performance comparison under different optimization strategies. (**a**) Fitness convergence behavior under different optimization strategies; (**b**) Efficiency-to-cost ratios of typical scenarios under different optimization strategies.

As shown in Figure 10b, the complete optimization scheme achieves higher efficiency–cost ratios across multiple representative scenarios. In particular, in urban high-interaction scenarios such as Scenarios 33 and 47, the improvement exceeds 25%, demonstrating the strategy's advantage in resource allocation under high-complexity environments.

4.2. Comparison of Efficiency–Cost Ratio

As shown in Figure 11, the complete optimization scheme achieves a higher efficiency–cost ratio across multiple typical scenarios. In particular, for urban interaction-intensive scenarios (such as Scenario 33 and Scenario 47), GA-SA-K Means improves the efficiency–cost ratio by over 25% compared with the baseline GA, and also demonstrates clear advantages over the single-mechanism variants GA-SA and GA-K Means. This indicates that the complete scheme provides superior resource allocation capability in high-complexity environments.

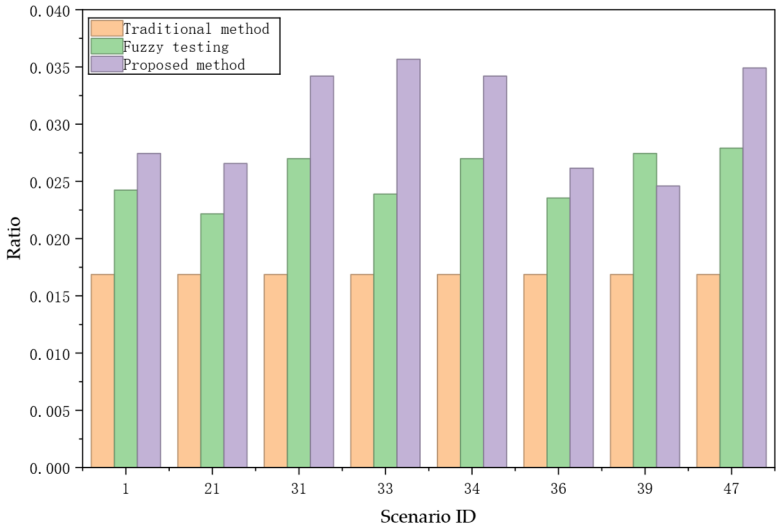


Figure 11. Test Efficiency and Cost Proportion.

Specifically, in high-interaction urban scenarios, the improvement in efficiency–cost ratio is most pronounced. Scenarios 33 and 47 both represent urban intersection conditions, featuring multiple motor vehicles, pedestrians, and potential occluding objects. Clear conflict points exist between vehicle trajectories, resulting in high levels of complexity and risk.

In these scenarios, the combinatorial testing strategy adaptively increases the proportion of open-road testing based on quantified complexity and risk, allocating more testing resources to real traffic environments to enhance the exposure probability of potential failure modes. Simultaneously, by jointly optimizing the fuzzy rule base and the testing ratios, the number of proving-ground tests in low-value conditions is reduced, thereby significantly improving the efficiency–cost ratio without increasing the overall testing budget.

In contrast, in scenarios such as 21 and 36, which represent low-speed or proving-ground conditions with relatively high traffic flow, the cost per test is low, but the risk exposure per test is limited. In these scenarios, the combinatorial testing strategy tends to execute test cases in bulk within proving grounds, allocating open-road tests only as necessary to ensure overall risk exposure while controlling the additional cost associated with open-road testing. The improvement in efficiency–cost ratio in these scenarios is lower than in high-interaction urban scenarios, but overall remains superior to the baseline, indicating that the strategy can achieve differentiated resource allocation according to scenario complexity and risk characteristics.

Overall, across the eight selected representative parameterized scenarios, the combinatorial testing strategy achieves higher efficiency–cost ratios than the unoptimized and single-mechanism baseline schemes in highway, urban, and proving-ground environments. This demonstrates that the model can adaptively adjust the proportion of proving-ground and open-road testing based on scenario complexity and risk levels, balancing risk exposure capability and cost control under a fixed testing budget, and providing a stable advantage in resource allocation.

4.3. Sensitivity Analysis with Respect to Expert-Defined Parameters

4.3.1. Experimental Setup

In the proposed fuzzy inference framework, all fuzzy rules are assigned equal activation weights and remain active during the inference process. Therefore, no relative importance weighting is introduced among rules. Under this setting, expert-defined parameters mainly refer to the initial parameters of the membership functions specified based on domain knowledge prior to optimization. The expert-defined parameters considered in this study are summarized in Table 7.

Table 7. Expert-defined parameters involved in the sensitivity analysis.

Parameter Category	Parameter Description	Symbol	Reference Value Source	Perturbed in Analysis
Membership function	Center of Gaussian membership function	μ	Expert-defined initialization	Yes
Membership function	Width of Gaussian membership function	σ	Expert-defined initialization	Yes
Fuzzy rule base	Rule activation status	–	Structural design always active	No

To examine the sensitivity of the proposed framework to variations in expert-defined parameters, a controlled perturbation analysis was conducted on the membership function parameters listed in Table 7. Specifically, the centers and widths of the Gaussian membership functions were independently perturbed within a $\pm 10\%$ range around their reference values. During the perturbation process, the linguistic partitioning of input variables and the semantic structure of the fuzzy rule base were preserved to ensure consistency in rule interpretation.

For each perturbed parameter configuration, the testing environment allocation ratios were recalculated using the same set of scenario inputs as those employed in the main experiments. To isolate the effect of parameter perturbations, the analysis was conducted using Scenario 47, which is adopted as a representative case in the main text. The sensitivity analysis was carried out under identical experimental settings to maintain consistency with the primary experiments.

4.3.2. Results and Analysis

The results indicate that perturbations in the membership function parameters lead to observable numerical variations in the testing environment allocation ratios, as illustrated in Figures 12 and 13. However, the overall allocation patterns and decision trends remain consistent across different parameter configurations.

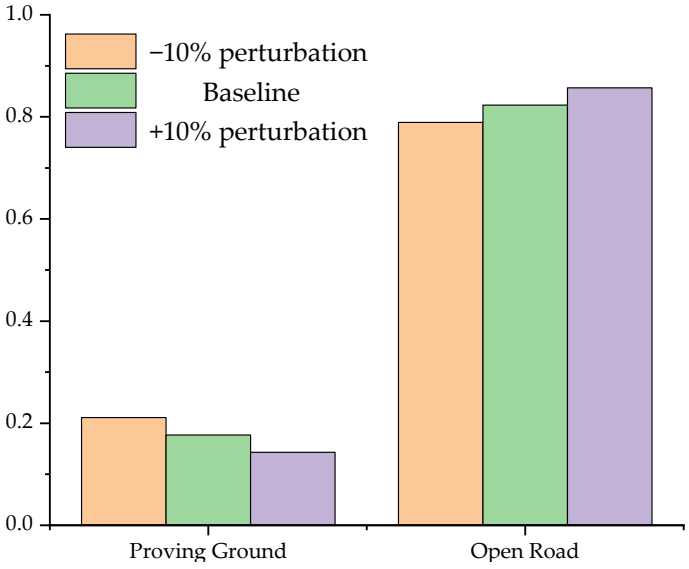


Figure 12. Sensitivity of testing environment allocation ratios to combined ±10% perturbations of membership function centers and widths for Scenario 47.

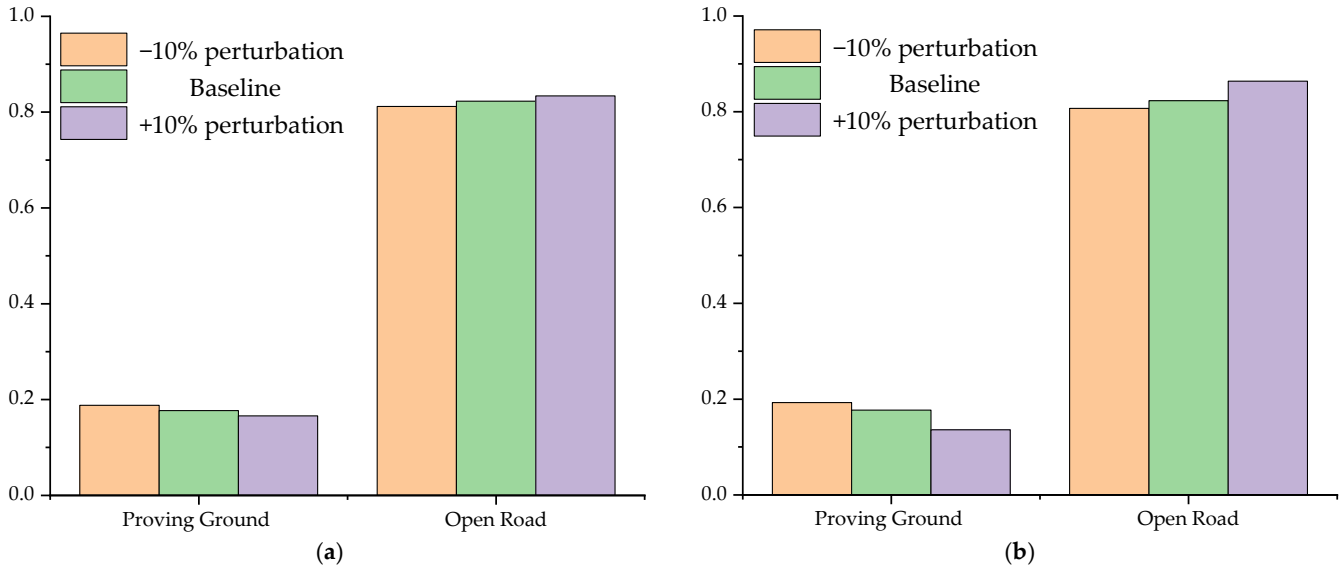


Figure 13. Sensitivity of testing environment allocation ratios to one-factor-at-a-time perturbations of membership function parameters for Scenario 47: (**a**) center perturbation only; (**b**) width perturbation only.

Specifically, scenarios characterized by higher risk levels consistently result in increased proportions of closed proving-ground testing, while scenarios with higher complexity levels continue to exhibit a greater reliance on open-road testing. Although minor fluctuations in allocation ratios are observed under parameter perturbations, the relative ordering of testing preferences across different environments is preserved.

These observations suggest that the testing environment allocation decisions produced by the proposed framework are stable with respect to reasonable variations in expert-defined membership function parameters. The sensitivity analysis confirms that the resulting allocation behavior follows consistent scenario-driven trends rather than being driven by specific parameter settings.

Overall, the sensitivity analysis results indicate that, under reasonable perturbations of the membership function parameters while preserving the semantic structure of the fuzzy rule base, only limited numerical variations are observed in the testing environment allocation ratios, whereas the overall allocation patterns and decision trends remain unchanged. Under both combined and one-factor-at-a-time perturbation settings, the relative ordering of testing environment preferences is consistently preserved. These results demonstrate that the proposed testing environment allocation mechanism exhibits stable mapping behavior with respect to expert-defined membership function parameters, and that the resulting allocation decisions are primarily driven by scenario characteristics rather than specific parameter settings.

4.4. Interpretability Analysis of the Optimized Fuzzy Rules

To further examine the interpretability of the proposed framework, this section analyzes the fuzzy rule base before and after optimization by visualizing the corresponding rule surfaces. The fuzzy rule surface provides an intuitive representation of how testing environment allocation responds to variations in scenario risk and complexity, thereby offering a direct means to assess whether the optimization process preserves the semantic structure of the original rules.

Figure 14 presents the fuzzy rule-base heatmaps before and after optimization, where Figure 14a shows the initial rule-base heatmap and Figure 14b illustrates the optimized one. As observed, the optimized rule-base heatmap preserves the overall monotonic trend with respect to scenario risk and complexity. Specifically, higher risk levels consistently correspond to increased allocations toward closed proving-ground testing, while higher complexity levels continue to exhibit a stronger preference for open-road testing.

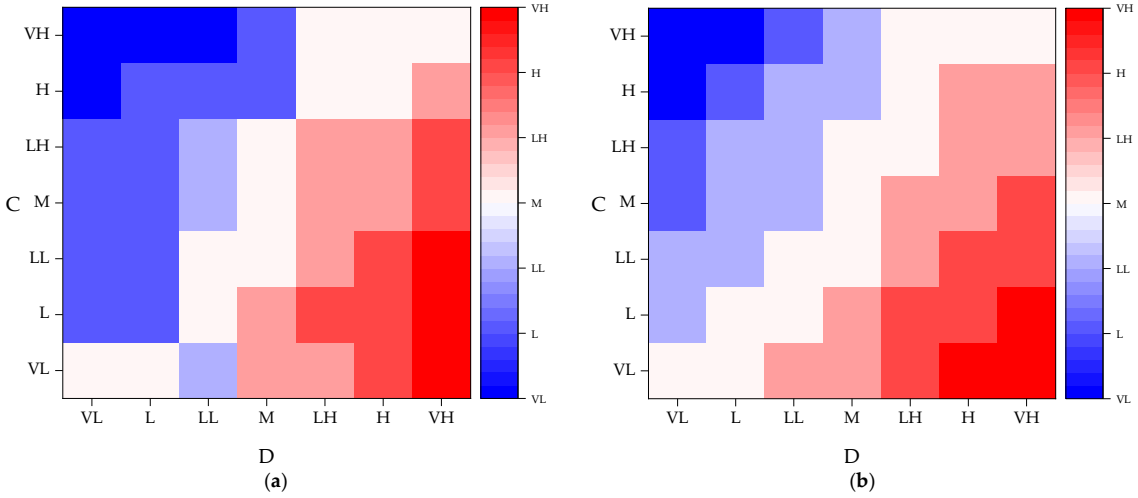


Figure 14. Comparison of fuzzy rule-base heatmaps before and after optimization: (**a**) initial rule-base heatmap; (**b**) optimized rule-base heatmap.

Compared with Figure 14a, the optimized rule-base heatmap in Figure 14b exhibits smoother transitions in regions associated with moderate risk and complexity, while more concentrated responses are observed under extreme scenario conditions. No structural inversions or counterintuitive patterns are introduced during optimization, and the relative ordering of testing environment preferences remains unchanged. These results indicate that the evolutionary optimization refines the inference behavior through parameter-level adjustments without altering the semantic structure of the original rule base.

5. Conclusions

This study proposes a fuzzy-system-based combinatorial testing strategy for automated driving systems, in which an improved genetic algorithm is integrated to optimize the allocation ratio between proving-ground testing and open-road testing. By jointly modeling scenario complexity and risk within a unified fuzzy inference framework, the proposed approach enables quantitative characterization of scenario-level testing requirements and supports interpretable testing resource allocation across heterogeneous driving environments. On this basis, the fuzzy rule base is further optimized using an improved genetic algorithm, allowing the allocation strategy to adapt to variations in scenario distributions while preserving rule-level interpretability. The framework incorporates multiple traffic interaction indicators, including information entropy, angular factors, relative speed, and TTC, to construct a conflict risk assessment model that does not rely on high-precision measurements, thereby extending its applicability to nonlinear and high-dimensional traffic conditions. Combined with reproducible scenario construction based on accident samples and naturalistic driving trajectories, and integrated with cost–efficiency modeling, the proposed strategy forms a closed-loop testing framework that supports quantitative evaluation and feedback of testing performance. Experimental results demonstrate that the optimized strategy achieves higher resource utilization efficiency and improved testing economy compared with traditional unoptimized approaches, particularly in scenarios characterized by high complexity and intensive interactions.

The proposed framework is developed and validated using traffic data collected from Chinese road environments, and its current scope is therefore aligned with the corresponding traffic characteristics. Nevertheless, the underlying modeling rationale—namely, the abstraction of scenario complexity and risk and the fuzzy inference-based testing environment allocation mechanism—is not inherently restricted to a specific region. For applications in other traffic systems, the overall framework can be retained while the statistical ranges and parameter distributions of scenario attributes are recalibrated using local data. Future work may further explore automatic scenario generation and semantic reconstruction methods to efficiently model both typical and edge-case traffic situations, as well as the integration of multi-source heterogeneous data such as V2X communication, infrastructure information, and sensor observations. In addition, advanced optimization methods, including reinforcement learning and evolutionary game-theoretic approaches, could be incorporated to systematically compare testing strategies under equivalent resource constraints, further improving the generality and adaptability of combinatorial testing frameworks for automated driving systems.

Author Contributions: Conceptualization, B.M. and Y.H.; methodology, B.M.; validation, B.M.; formal analysis, B.M.; investigation, B.M.; resources, Y.H.; data curation, B.M.; writing—original draft preparation, B.M.; writing—review and editing, X.Z. and B.M.; visualization, B.M.; supervision, X.Z. and H.S.; project administration, X.Z. and L.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key R&D Program of China (2022YFB2503404).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study are not publicly available at this time due to restrictions related to the collaborating enterprise's data management policies and the planned use of these data in subsequent research and publications. Data may be made available by the corresponding author upon reasonable request and with permission from the enterprise.

Acknowledgments: This article is grateful to the China Automotive Technology and Research Center for providing data support.

Conflicts of Interest: Authors H.S. and Y.H. were employed by China Automotive Technology and Research Center Co. Ltd. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Abdelkader, G.; Elgazzar, K.; Khamis, A. Connected Vehicles: Technology Review, State of the Art, Challenges and Opportunities. *Sensors* **2021**, *21*, 7712. [CrossRef] [PubMed]
2. Autonomous Driving Working Group. *Simulation, Testing, Verification, and Validation (STV^2) of Autonomous Driving*; IEEE Standards Association: Piscataway, NJ, USA, 2024.
3. Ploeg, J.; de Gelder, E.; Slavík, M.; Querner, E.; Webster, T.; de Boer, N. Scenario-Based Safety Assessment Framework for Automated Vehicles. *arXiv* **2021**, arXiv:2112.09366. [CrossRef]
4. *GB/T 44719-2024*; Intelligent and Connected Vehicle—Methods and Requirements of Road Test for Automated Driving Functions. China National Standard: Beijing, China, 2024.
5. Zhong, Z.; Tang, Y.; Zhou, Y.; Neves, V.D.O.; Liu, Y.; Ray, B. A survey on scenario-based testing for automated driving systems in high-fidelity simulation. *arXiv* **2021**, arXiv:2112.00964.
6. Xu, Z.; Wang, X.; Wang, X.; Zheng, N. Safety Validation for Connected Autonomous Vehicles Using Large-Scale Testing Tracks in High-Fidelity Simulation Environment. *Accid. Anal. Prev.* **2025**, *215*, 108011. [CrossRef] [PubMed]
7. Duan, X.; Yang, Y.; Tian, D.; Wang, Y.; Li, T. A V2X Communication System and Its Performance Evaluation Test Bed. In Proceedings of the 2014 IEEE 6th International Symposium on Wireless Vehicular Communications (WiVeC 2014), Vancouver, BC, Canada, 14–15 September 2014; pp. 1–2.
8. Shi, M.; Lu, C.; Zhang, Y.; Yao, D. DSRC and LTE-V Communication Performance Evaluation and Improvement Based on Typical V2X Application at Intersection. In Proceedings of the 2017 Chinese Automation Congress (CAC), Jinan, China, 20–22 October 2017; pp. 556–561.
9. Maghsoumi, H.; Fallah, Y. ConvoyNext: A Scalable Testbed Platform for Cooperative Autonomous Vehicle Systems. *arXiv* **2025**, arXiv:2505.17275. [CrossRef]
10. Bein, T.; Atzrodt, H.; Bartolozzi, R.; Kupjetz, S.; Millitzer, J.; Nuffer, J.; Rauschenbach, M.; Stoll, G. Verification and Validation of Automated Driving Systems Utilizing Probabilistic FMEA and Simulation Approaches. *Transp. Res. Procedia* **2023**, *72*, 470–477. [CrossRef]
11. Zhang, S.; Zhao, C.; Zhang, Z.; Lv, Y. Driving Simulator Validation Studies: A Systematic Review. *Simul. Model. Pract. Theory* **2025**, *138*, 103020. [CrossRef]
12. Alemayehu, H.; Sargolzaei, A. Testing and Verification of Connected and Autonomous Vehicles: A Review. *Electronics* **2025**, *14*, 600. [CrossRef]
13. Ren, K.; Yang, J.; Lu, Q.; Zhang, Y.; Hu, J.; Feng, S. Intelligent Testing Environment Generation for Autonomous Vehicles with Implicit Distributions of Traffic Behaviors. *Transp. Res. Part C Emerg. Technol.* **2025**, *174*, 105106. [CrossRef]
14. Feng, S.; Feng, Y.; Yan, X.; Shen, S.; Xu, S.; Liu, H.X. Safety Assessment of Highly Automated Driving Systems in Test Tracks: A New Framework. *Accid. Anal. Prev.* **2020**, *144*, 105664. [CrossRef] [PubMed]
15. Feng, S.; Yan, X.; Sun, H.; Feng, Y.; Liu, H.X. Intelligent Driving Intelligence Test for Autonomous Vehicles with Naturalistic and Adversarial Environment. *Nat. Commun.* **2021**, *12*, 748. [CrossRef] [PubMed]
16. Fremont, D.J.; Kim, E.; Pant, Y.V.; Seshia, S.A.; Acharya, A.; Bruso, X.; Wells, P.; Lemke, S.; Lu, Q.; Mehta, S. Formal Scenario-Based Testing of Autonomous Vehicles: From Simulation to the Real World. In Proceedings of the 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), Rhodes, Greece, 20–23 September 2020; pp. 1–8.
17. Alghodhaifi, H.; Lakshmanan, S. Autonomous vehicle evaluation: A comprehensive survey on modeling and simulation approaches. *IEEE Access* **2021**, *9*, 151531–151566. [CrossRef]
18. Tang, S.; Zhang, Z.; Zhang, Y.; Zhou, J.; Guo, Y.; Liu, S.; Liu, Y. A survey on automated driving system testing: Landscapes and trends. *ACM Trans. Softw. Eng. Methodol.* **2023**, *32*, 1–62. [CrossRef]

19. Eisemann, L.; Fehling-Kaschek, M.; Forkert, S.; Forster, A.; Gommel, H.; Guenther, S.; Hammer, S.; Hermann, D.; Klemp, M.; Lickert, B.; et al. A Joint Approach Towards Data-Driven Virtual Testing for Automated Driving: The AVEAS Project. *arXiv* **2024**, arXiv:2405.06286. [CrossRef]
20. Betschinske, D.; Schrimpf, M.; Peters, S.; Klonecki, K.; Karch, J.P.; Lippert, M. Towards More Efficient Quantitative Safety Validation of Residual Risk for Assisted and Automated Driving. *arXiv* **2025**, arXiv:2506.10363. [CrossRef]
21. Wei, Z.; Zhou, H.; Zhou, R. Risk and Complexity Assessment of Autonomous Vehicle Testing Scenarios. *Appl. Sci.* **2024**, *14*, 9866. [CrossRef]
22. Ponn, T.; Breitfuß, M.; Yu, X.; Diermeyer, F. Identification of Challenging Highway-Scenarios for the Safety Validation of Automated Vehicles Based on Real Driving Data. In Proceedings of the 15th International Conference on Ecological Vehicles and Renewable Energies (EVER 2020), Monte-Carlo, Monaco, 10–12 September 2020; IEEE: Piscataway, NJ, USA; pp. 1–10.
23. Ren, D.; Huang, H.; Li, Y.; Jin, J. High-Risk Test Scenario Generation for Autonomous Vehicles at Roundabouts Using Naturalistic Driving Data. *Appl. Sci.* **2025**, *15*, 4505. [CrossRef]
24. Menzel, T.; Bagschik, G.; Maurer, M. Scenarios for Development, Test and Validation of Automated Vehicles. In Proceedings of the 2018 IEEE Intelligent Vehicles Symposium (IV), Changshu, China, 26–30 June 2018; pp. 1821–1827.
25. Wu, B.; Ren, H.; Zheng, L.; Zhu, X.; Ma, Z. Complex Scenario Construction Method for Navigation Pilot Based on Natural Driving Behaviour. *J. South China Univ. Technol. (Nat. Sci. Ed.)* **2025**, *53*, 38–47.
26. Hayward, J.C. *Near Miss Determination through Use of a Scale of Danger*; Highway Research Record; Highway Research Board, Transportation Research Board (TRB): Washington, DC, USA, 1972.
27. Mendel, J.M. *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*; Prentice Hall: Upper Saddle River, NJ, USA, 2001.
28. Cord, O. *Genetic Fuzzy Systems: Evolutionary Tuning and Learning of Fuzzy Knowledge Bases*; World Scientific: Singapore, 2001; Volume 19.
29. Karr, C. Genetic Algorithms for Fuzzy Controllers. *AI Expert* **1991**, *6*, 26–33.
30. Nellen, J.; Wolters, B.; Netz, L.; Geulen, S.; Ábrahám, E. A Genetic Algorithm-Based Control Strategy for the Energy Management Problem in PHEVs. In Proceedings of the GCAI 2015, Tbilisi, Georgia, 16–19 October 2015; pp. 196–214.
31. Arief, M.; Glynn, P.; Zhao, D. An Accelerated Approach to Safely and Efficiently Test Pre-Production Autonomous Vehicles on Public Streets. In Proceedings of the 2018 IEEE 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018; pp. 2006–2011.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.