World Journal of Engineering Research
Print ISSN: 2959-9865
Online ISSN: 2959-9873
DOI: https://doi.org/10.61784/wjer3090

THE KEY TECHNOLOGIES FOR RISK IDENTIFICATION OF "TWO PASSENGER AND ONE HAZARDOUS" VEHICLES BASED ON MULTI-SOURCE DATA FUSION

HengBo Zhang[1*], WanMiao Yu[2], Long Peng[2], ChunFu Jia[2], ZhiGang Wei[3], Miao Li[1]
[1]*Beijing CCCC Intelligent Transportation System Technology Co., Ltd., Beijing 100088, China.*
[2]*Jilin Provincial Transportation Planning and Design Institute, Changchun 130021, Jilin, China.*
[3]*Jilin Provincial High-grade Highway Construction Bureau, Changchun 130100, Jilin, China.*
[*]*Corresponding Author: HengBo Zhang*

Abstract: "Two passenger and one hazardous" (TPOH) vehicles are primary targets of road traffic safety supervision. Jilin Province, located in the northeastern seasonal frozen region of China, experiences severe winter conditions, including icy and snow-covered roads and extremely low temperatures, which substantially increase the operational safety risks of these vehicles.This study proposes a risk identification framework based on multi-source data fusion. The framework integrates GPS positioning data, on-board OBD data, meteorological information, and road infrastructure data to construct a multidimensional risk identification model. The model captures five categories of abnormal driving behaviors: speeding, harsh acceleration, harsh braking, vehicle vibration, and abrupt lane changes.An improved sliding-window feature extraction method and a hybrid XGBoost–LSTM classification model are employed to achieve accurate vehicle risk identification under complex seasonal frozen conditions. In addition, a system management and control platform is developed to provide decision-support tools for the transportation authorities of Jilin Province.
Keywords: Two Passenger and One Hazardous(TPOH); Multi-source data fusion; Risk identification; Seasonal frozen region, XGBoost-LSTM

1 INTRODUCTION

"Two passenger and one hazardous" (TPOH) vehicles are high-risk carriers in the road transportation sector. Due to their large passenger capacity and the hazardous nature of the goods transported, their operational safety is directly linked to public safety and ecological security.Jilin Province is a typical seasonal frozen region, where winter temperatures can fall below −40 °C. These extreme climatic conditions frequently lead to road surface icing, snow accumulation, and freeze–thaw damage. Such conditions increase braking distance and reduce vehicle stability. In addition, low temperatures degrade mechanical performance and impair drivers' physiological responsiveness. The combined effects of these factors substantially amplify the operational risks of TPOH vehicles [1].Traditional monitoring approaches based on a single data source are insufficient to comprehensively capture the coupled risk characteristics in the complex environment of seasonal frozen regions. As a result, they fail to achieve refined identification of vehicle operational risks. Therefore, targeted research based on multi-source heterogeneous data fusion is urgently needed.

Multi-source data fusion technology provides a novel methodological pathway for road traffic safety risk analysis and has become a research hotspot in this field. However, existing studies primarily focus on conventional climatic and road conditions. Systematic investigations into multidimensional and coupled risk identification of "Two passenger and one hazardous" (TPOH) vehicles under the unique conditions of seasonal frozen regions remain limited [2]. Moreover, a risk identification framework and management strategy tailored to the environmental characteristics of seasonal frozen regions has yet to be established.In response to this gap, this study takes the seasonal frozen region of Jilin Province as the study area and focuses on the operational risk identification and management requirements of TPOH vehicles. A key technical framework for risk identification based on multi-source heterogeneous data fusion is proposed, and a supporting safety management platform is developed. This work aims to address the deficiencies in risk identification research under seasonal frozen conditions and to provide theoretical support and technical assurance for the refined management of road transportation safety in such regions.

2 OVERALL FRAMEWORK

This study adopts a four-layer architecture consisting of a data acquisition layer, data processing layer, model analysis layer, and application service layer. The overall framework is illustrated in Figure 1.

The data acquisition layer collects information from multiple sources, including GPS terminals, OBD interfaces, meteorological stations, road databases, and video surveillance systems.The data processing layer performs data cleaning, temporal and spatial alignment, feature engineering, and parameter calibration tailored to seasonal frozen conditions.The model analysis layer employs a hybrid XGBoost–LSTM model to identify abnormal driving

behaviors.The application service layer provides functions including real-time monitoring, early warning, report generation, and decision support.

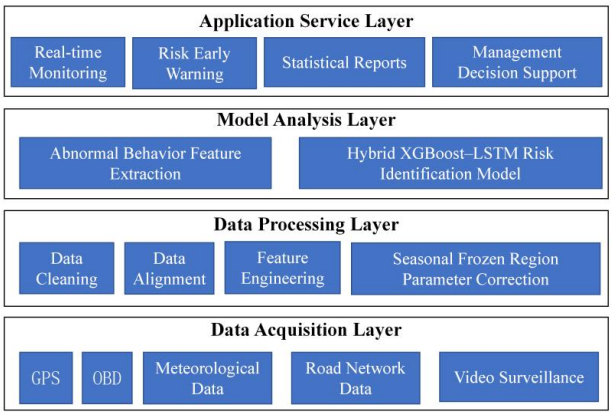

Figure 1 System Architecture Framework

3 DATA ANALYSIS AND PROCESSING

3.1 Multi-Source Data Acquisition

The data sources include four categories:
a. GPS positioning data were collected at a frequency of once every 10 seconds. These data were primarily used for vehicle trajectory tracking and speed monitoring, whereas instantaneous driving behaviors, such as abrupt lane changes, were identified using high-frequency OBD data. The GPS dataset includes fields such as latitude, longitude, speed, heading angle, and timestamp. It covers more than 3,200 “Two Passenger and One Hazardous” (TPOH) vehicles registered and operating in Jilin Province. The data collection period spanned from January to June 2025.
b. On-board OBD terminal data, collected at a frequency of once per second, containing dynamic parameters such as engine speed, vehicle speed, longitudinal and lateral acceleration, and steering wheel angle;
c. Meteorological data, sourced from 57 ground-based meteorological observation stations of the Jilin Provincial Meteorological Bureau, including observed data on temperature, snowfall, visibility, and wind speed and direction, with a temporal resolution of one hour;
d. Road infrastructure data, sourced from the GIS road network database of the provincial highway management department, containing static attribute information such as road grade, number of lanes, gradient, curve curvature, and speed limit signs.

3.2 Data Preprocessing

GPS data were first preprocessed by removing drift points, defined as records in which the single-step displacement exceeded the maximum physically plausible distance derived from the speed limit. Kalman filtering was then applied to smooth the trajectories and suppress signal jump noise [3].Missing OBD values were imputed using linear interpolation. Outliers exceeding physically reasonable thresholds (e.g., absolute acceleration greater than 15 m/s²) were removed.OBD data were downsampled using GPS timestamps as the temporal reference. During this process, spatial coordinates were assigned to each high-frequency OBD record.Specifically, for two adjacent GPS timestamps t_k and t_{k+1} ,cubic spline interpolation was performed on the latitude and longitude sequences. The estimated position at each OBD timestamp t_i was calculated using the temporal proportion factor $\tau=(t_i-t_k)/(t_{k+1}-t_k)$ .Slowly varying parameters, such as engine speed, were aggregated within a 10-second window by extracting their mean and extreme values as associated features. In contrast, transient parameters, including longitudinal and lateral acceleration, were retained at a one-second resolution without compression.Meteorological data were linked to trajectory points based on the nearest temporal matching principle. Road infrastructure attributes were matched according to spatial location.

3.3 Seasonal Frozen Region Environmental Correction Coefficient

This study introduces a comprehensive environmental correction coefficient, denoted as αenv, to dynamically adjust the risk thresholds:

$$\alpha_{env} = \omega_1 f(T) + \omega_2 g(\mu) + \omega_3 h(V) + \omega_4 p(S_n) \quad (1)$$

where $f(T)$ is the temperature decay function, whose value is greater than 1 when the temperature falls below 0°C, indicating that risk thresholds should be appropriately lowered under low-temperature conditions to more readily trigger alerts; $g(\mu)$ is the road surface friction coefficient impact function, where a lower friction coefficient yields a larger correction coefficient; $h(V)$ is the visibility impact function; and $p(S_n)$ is the snowfall intensity impact function. The weights are determined through the Analytic Hierarchy Process (AHP) combined with expert scoring as $\omega_1 = 0.25$, $\omega_2 =$

0.35, ω_3 = 0.20, and ω_4 = 0.20, where the road surface friction coefficient receives the highest weight, reflecting its critical influence on driving safety in the seasonal frozen region.

4 RISK IDENTIFICATION MODEL FOR TPOH VEHICLES BASED ON MULTI-SOURCE DATA

4.1 Definition of Abnormal Driving Behaviors and Feature Extraction

Abnormal driving behaviors are classified into five categories. The determination criteria are shown in Table 1.

Table 1 Classification and Determination Criteria of Abnormal Driving Behaviors

Behavior Type	Core Feature Parameters	Basic Determination Criteria	Seasonal Frozen Region Correction Rules
Speeding	Speed v, speed limit v_{lim}	$V > 1.1\ v_{lim}$	No exceeding of speed limit allowed on icy/snowy roads
Harsh Acceleration	Longitudinal acceleration a_x	$a_x > 2.5\ m/s^2$	Threshold reduced by 1.5 m/s^2 when $\mu < 0.3$
Harsh Braking	Longitudinal acceleration a_x	$a_x < 2.5\ m/s^2$	Threshold adjusted by $-2\ m/s^2$ when $\mu < 0.3$
Vehicle Vibration	Variance of vertical acceleration $\sigma_{a_z}^2$	Exceeding threshold σ_{th}	σ_{th} reduced by 20% during freeze–thaw cycles
Abrupt Lane Change	Lateral acceleration a_y, heading angle change rate φ	Both exceeding respective thresholds simultaneously	Thresholds reduced to 60% of original values on icy roads

An improved sliding window method is adopted to segment the time-series data [4], with a window length of 5 seconds and a step size of 1 second. Within each window, statistical features including mean, standard deviation, maximum, minimum, kurtosis, skewness, and the mean and standard deviation of first-order differences are extracted, ultimately forming a 32-dimensional fused feature vector. The environmental correction coefficient α_{env} is incorporated as an additional feature, enabling the model to perceive environmental changes in the seasonal frozen region.

4.2 XGBoost-LSTM Fusion Model

This study proposes a hybrid XGBoost–LSTM model that integrates efficient classification of static features with dynamic temporal modeling capabilities [5]. The model architecture is illustrated in Figure 2.

The XGBoost branch takes single-window features and outputs a six-class probability vector $P_{xgb} \in R^6$ (normal, speeding, harsh acceleration, harsh braking, vibration, and abrupt lane change). The LSTM branch takes a temporal sequence of 10 consecutive windows, processes it through a two-layer LSTM network (with 128 and 64 hidden units, respectively) to capture the temporal evolution patterns of driving behaviors, and outputs $P_{lstm} \in R^6$. The fusion probability is calculated as:

$$P_{final} = \lambda \cdot P_{xgb} + (1 - \lambda) \cdot P_{lstm} \quad (2)$$

The adaptive weight is $\lambda = 0.6 - 0.1\min(\alpha_{env}, 2.0)$, ensuring that λ fluctuates within the range of 0.4 to 0.6. When environmental conditions undergo drastic changes (i.e., α_{env} is large), the weight of the LSTM branch is increased to better capture dynamic trends. Conversely, when the environment is relatively stable, the weight of the XGBoost branch is increased to improve classification efficiency.

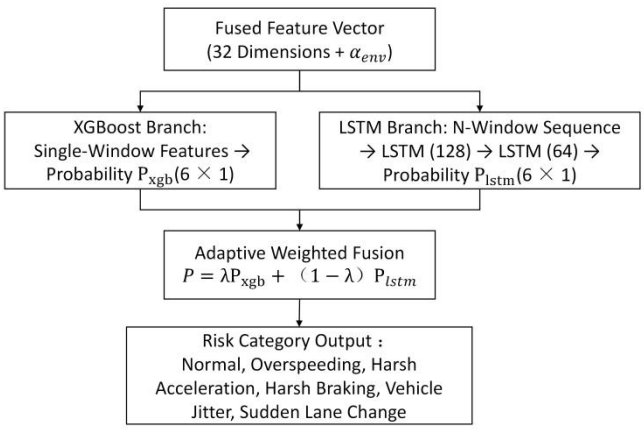

Figure 2 Structure of the XGBoost–LSTM Fusion Model

4.3 Model Validation

Data from January to March 2025 were used for model training, while data from April to June 2025 were reserved for testing. The dataset was further split into training, validation, and testing sets at a ratio of 70%, 15%, and 15%, respectively.In the XGBoost branch, the maximum tree depth was set to 8, the learning rate to 0.05, and the number of estimators to 500.The LSTM branch employed the Adam optimizer [6] with a learning rate of 0.001 and a batch size of 256. The model was trained for 100 epochs with an early stopping mechanism to prevent overfitting. The experimental results are presented in Table 2.

The hybrid model significantly outperforms the single models across all evaluation metrics. After incorporating the environmental correction mechanism, the accuracy increased from 90.1% to 92.6%, and the AUC improved from 0.948 to 0.967.Notably, the recall rates for harsh braking and abrupt lane changes exhibited the most substantial improvements, increasing by 4.8 and 5.2 percentage points, respectively. This finding is consistent with real-world conditions in seasonal frozen regions, where icy road surfaces increase the likelihood of braking instability and lane-deviation events.

A total of 12,846 test windows under icy conditions ($\mu < 0.3$) were further selected, and a normalized confusion matrix was computed, as shown in Table 3. The results indicate a noticeable mutual misclassification between “vehicle vibration” and “hard braking,” with rates of 5.3% and 4.7%, respectively. This confusion is primarily attributed to the similarity in vertical acceleration variance features between high-frequency vibrations induced by ABS activation during hard braking on icy roads and those caused by pavement roughness under freeze–thaw conditions. In contrast, the corresponding misclassification rates under non-icy conditions were only 1.8% and 1.5%.

Table 2 Model Performance Comparison (Test Set)

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	AUC
Single XGBoost	87.3	85.6	84.2	84.9	0.921
Single LSTM	85.8	84.1	86.5	85.3	0.913
Random Forest	84.6	83.2	82.8	83.0	0.905
XGBoost-LSTM(w/o correction)	90.1	88.7	89.3	89.0	0.948
XGBoost-LSTM(with correction)	92.6	91.2	91.8	91.5	0.967

Table 3 Normalized Confusion Matrix under Icy Conditions (%)

Actual \ Predicted	Normal Driving	Speeding	Harsh Acceleration	Harsh Braking	vehicle vibration	Abrupt Lane Change
Normal Driving	94.1	1.8	1.2	1.5	0.9	0.5
Speeding	2.3	93.5	0.8	1.1	0.6	1.7
Harsh Acceleration	1.9	0.7	90.4	3.8	2.1	1.1
Harsh Braking	1.6	0.5	3.2	88.6	4.7	1.4
vehicle vibration	1.2	0.4	1.8	5.3	89.2	2.1
Abrupt Lane Change	0.8	1.5	0.9	1.7	2.3	92.8

Note: The diagonal cells (blue background) represent the correct classification rates. The cells highlighted in red indicate mutual misclassification between “Vehicle Vibration” and “Hard Braking.”

5 DESIGN OF THE TPOH VEHICLE SYSTEM MANAGEMENT AND CONTROL PLATFORM

Based on the aforementioned model, a management and control platform is designed for the transportation management authorities and transport enterprises of Jilin Province. The platform adopts a B/S architecture, with the back end based on the Spring Cloud microservices framework [7] and the front end implemented using Vue.js for visual interaction. A hybrid storage solution comprising MySQL + Redis + InfluxDB is employed [8], where InfluxDB is used for high-frequency time-series data.The platform architecture is illustrated in Figure 3.

The core functional modules include five aspects:

a. **Real-time monitoring module**: Based on a GIS-enabled digital map, this module enables system-wide vehicle location tracking and status visualization. Risk levels are indicated using color-coded markers (green for normal, yellow for low risk, orange for medium risk, and red for high risk), with real-time meteorological and road surface condition layers overlaid for seasonal frozen regions [9].

b. **Risk early warning module**: Upon detection of abnormal behaviors, three levels of alerts (general, important, and urgent) are synchronously issued through multiple channels, including SMS notifications, pop-up messages, and in-vehicle voice prompts. Under severe weather conditions in seasonal frozen regions, the warning thresholds are automatically lowered.

c. **Statistical analysis module**: This module performs multidimensional aggregation and trend analysis based on time, region, vehicle type, and behavior category. Risk distributions are visualized using line charts, heat maps, and other graphical representations, with particular emphasis on comparing seasonal frozen and non-frozen periods.
d. **Enterprise evaluation module**: Monthly safety scores and rankings are generated based on comprehensive indicators, including abnormal behavior frequency, warning response time, and training completion rate. The evaluation results are linked to the annual review of operational qualifications, thereby establishing an incentive and accountability mechanism [10].
e. **Emergency response module:** When a critical event is triggered, the system automatically locks the vehicle location, notifies nearby rescue resources, and archives the entire response process. For hazardous materials vehicles, cargo information and emergency handling guidelines can be rapidly retrieved.

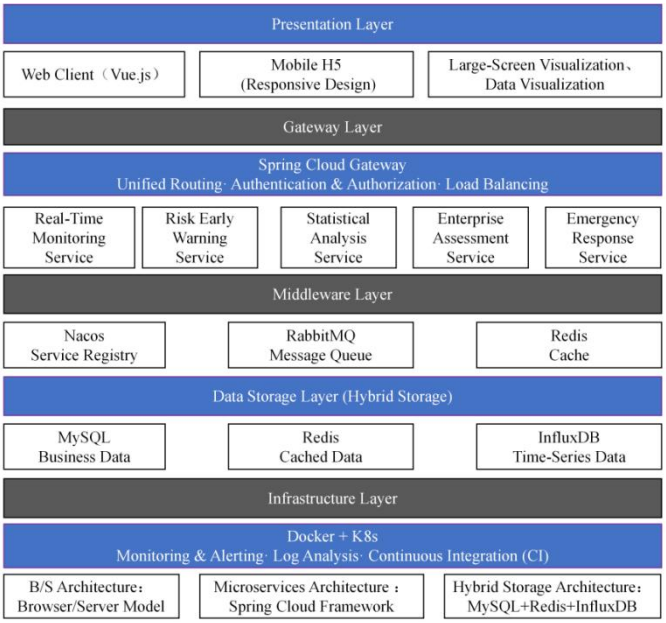

Figure 3 Architecture of the Control and Management Platform

6 CONCLUSIONS

To address the safety supervision needs of “Two Passenger and One Hazardous” (TPOH) vehicles in the seasonal frozen regions of Jilin Province, this study develops a multi-source data fusion framework integrating GPS, OBD, meteorological, and road infrastructure data. It further establishes an environmental factor system and a comprehensive correction coefficient calculation method tailored to seasonal frozen regions.Five categories of abnormal driving behaviors—speeding, rapid acceleration, hard braking, vehicle vibration, and abrupt lane change—are characterized through dedicated feature representations and correction rules adapted to seasonal frozen conditions. Based on these features, an XGBoost–LSTM hybrid model is proposed for risk identification.In addition, a system management platform was developed with functions including real-time monitoring, risk warning, statistical analysis, enterprise assessment, and emergency response, providing a comprehensive technical solution for transportation safety management in Jilin Province.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Li C, Yang Y, Cao G. Quadruple-U Auxiliary Structure-Based Receiving Coil Positioning System for Electric Vehicle Wireless Charger. World Electric Vehicle Journal, 2023, 14(5): 115.
[2] Ma W, Yuan J, An K, et al. Route flow estimation based on the fusion of probe vehicle trajectory and automated vehicle identification data.Transportation Research Part C:Emerging Technologies, 2022, 144: 103907.
[3] Qi X, Ji Y, Li W, et al. Vehicle trajectory reconstruction on urban traffic network using automatic license plate recognition data. IEEE Access, 2021, 9: 49110-20. DOI: 10.1109/ACCESS.2021.3068866.
[4] Yao Z, Liu M, JiangY, et al. Trajectory reconstruction for mixed traffic flow with regular,connected, and connected automated vehicles on freeway. IET Intelligent Transport Systems, 2024, 18(3): 450-66. DOI: 10.1049/itr2.12457.
[5] Huang SE, Feng Y, Liu HX. A data-driven method for falsified vehicle trajectory identification by anomaly detection. Transportation research part C: emerging technologies, 2021, 128: 103196.

[6] He L, Niu X, Chen T, et al. Spatio-temporal trajectory anomaly detection based on common sub-sequence. Applied Intelligence, 2022: 1-23. DOI: 10.1007/s10489-022-04330-3.
[7] Li C, Feng G, Li Y, et al. DiffTAD: Denoising diffusion probabilistic models for vehicle trajectory anomaly detection. Knowledge-Based Systems, 2024, 286: 111387. DOI: 10.1016/j.knosys.2024.111387.
[8] Zhang Y, He Y, Zhang L. Recognition method of abnormal driving behavior using the bidirectional gated recurrent unit and convolutional neural network. Physica A: Statistical Mechanics and its Applications. 2023, 609: 128317. DOI: 10.1016/j.physa.2022.128317.
[9] Ma Y, Xie Z, Chen S, et al. Real-time detection of abnormal driving behavior based on long short-term memory network and regression residuals. Transportation research part C:emerging technologies, 2023, 146: 103983. DOI: 10.1016/j.trc.2022.103983.
[10] Shi Y, Wang D, Tang J B, et al. Detecting spatiotemporal extents of traffic congestion: A density-based moving object clustering approach. International Journal of Geographical Information Science, 2021, 35(7): 1449-1473. DOI: 10.1080/13658816.2020.1862853.