Journal of Applied Automation Technologies

Research Paper

Real-Time Traffic Accident Risk Assessment Based on Multimodal Sensor Data Fusion

Nikodem Baran[1, *] and Mateusz Robert Cichy[1]

[1] Faculty of Mechatronics and Automatics, Lodz University of Technology, Lodz, 90-924, Poland
*Corresponding author: nikodem.b@p.lodz.pl

Abstract. For intelligent transportation systems, real-time traffic accident risk assessment is crucial, particularly in complicated urban road circumstances where vehicle trajectories, roadside perception, weather, speed, and traffic flow interact dynamically. A multimodal sensor data fusion framework for real-time accident risk assessment is proposed in this research. The technique creates a single spatiotemporal representation by combining camera-based visual features, radar motion data, vehicle trajectory data, traffic flow statistics, and environmental sensing factors. Short-term motion conflict, local traffic instability, and cross-modal risk consistency are captured by a lightweight fusion network. According to experimental results, the suggested approach outperforms single-modal and early-fusion baselines with an accuracy of 94.1%, an F1-score of 92.8%, an AUC of 96.2%, and an average inference latency of 31.6 ms. The findings show that while preserving real-time applicability for intelligent traffic monitoring systems, multimodal fusion can enhance accident risk discrimination.

Keywords: *Multimodal Sensor Data Fusion, Traffic Accident Risk Assessment, Real-Time Prediction, Intelligent Transportation Systems, Spatiotemporal Feature Learning*

Received on 25 August 2022, Accepted on 03 April 2023, Published on 16 April 2023

Introduction

Ensuring road traffic safety has emerged as a major engineering challenge with the advent of intelligent transportation systems, connected vehicles, and urban traffic management. Cameras, millimeter-wave radar, loop detectors, roadside units, floating vehicle trajectories, weather sensors, and signal controllers are all installed on New Roads. Data on vehicle speed, lane-change behavior, traffic density, acceleration fluctuations, queue propagation, visibility, rainfall, and local congestion are all continuously gathered by the sensing sources. Unstable traffic circumstances, aggressive driving, short following distances, incorrect lane changes, poor vision, and slow driver reactions all contribute to accident risk, which is rarely caused by a single factor. Data-driven models have recently been used in traffic accident risk prediction research to find non-linear accident precursors in vehicular networks and traffic data [1]. Additionally, deep learning has been applied to increase the accuracy of crash severity prediction and uncover hidden risk patterns of various transportation parameters [2]. The focus of accident prediction research has shifted from retrospective statistical explanation to proactive prevention, with the aim now being to detect unstable traffic conditions before to an accident [3]. Dangerous driving practices and the local road environment both influence the likelihood of an accident, according to city-wide fusion models [4]. Research on machine learning-based traffic risk has also discovered that heterogeneous features improve warning reliability after processing in a single analytical framework [5].

Due to highly diverse, noisy, asynchronous, and geographically correlated transportation data, real-time accident-risk assessment remains a challenge despite the increasing complexity of traffic-sensing systems. Trajectory data describes continuous motion; radar signals provide relative speed and distance; traffic flow statistics reflect network-wide instability; camera images display lane occupation and visual conflicts; and

https://doi.org/10.64972/jaat.2023v1.287p9e:118-133

environmental sensors gather external risk conditions. Occlusion, sensor interference, or inclement weather might cause a single mode to fail, and basic early fusion can produce duplicate or contradicting data. The development of an urban traffic safety system should incorporate road, vehicle, and environment data through numerous sensors rather than employing a single sensor, according to several sources of accident prevention research [6]. The main causes of accidents differ by intersection, merging area, and pedestrian conflict zone, according to explainable accident risk models [7]. By combining traffic, weather, and other elements, several data sources have been leveraged to increase the accuracy of risk prediction [8]. Additionally, risk cues are weak in individual dimensions but significant following cross-feature coupling, according to multiple feature interaction models [9]. A short headway and relative speed must be taken into account jointly for early warning, according to proactive crash risk modeling in merging regions [10]. The need for temporal continuity and local road context is the main focus of research on spatiotemporal crash prediction [11]. The distribution of hazardous traffic contacts varies with space, according to pedestrian collision risk assessments based on POI and trip parameters [12]. Uncertainty estimation and sensor fusion are necessary for stable transportation monitoring, according to multi-camera and multi-sensor perception studies [13]. Because roadside devices typically cannot operate huge offline networks, compact perception models are also necessary [14]. Real-time safety applications require reliable output under sensor disruption and environmental changes, as demonstrated by driver warning systems based on several sensors [15].

Here, a novel approach to real-time traffic accident risk assessment based on multi-modal sensor data fusion is put forward. The goal is to develop a system that can display dangers in all-weather situations by integrating visual information, radar data, vehicle movement, traffic circumstances, etc. The three technical issues with the suggested approach are how to generate precise risk scores with minimal inference delay, how to extract discriminative spatiotemporal risk characteristics from various modalities, and how to align asynchronous sensor observations into a stable traffic state. The suggested framework is neither a single-sensor nor a static accident classification system; rather, it is for short-term risk discrimination under dynamic traffic situations. Reduce false alerts during bad weather, increase the precision of risk-level classification, and provide a deployable inference structure for edge-side intelligent transportation systems and roadside monitoring units. Delayed accident labels were not used in this study because of the engineering presumption that accident risk should be evaluated under changing traffic conditions. Speed limit adjustments, traffic signal changes, lane warnings, and other human control actions can all be carried out with a notice that is several seconds in advance. As a result, the model should pick up on antecedent trends including decreased headways, erratic speed variations, inconsistent sensor data, and unstable lane occupancy. Therefore, a single sensor pipeline cannot satisfy this demand because these sensors cover diverse aspects of the risk process.

Related Work

Traffic Accident Risk Prediction Based on Sensor Data

Machine learning and deep learning techniques have replaced statistical crash-frequency models in the prediction of traffic accident risk. In the past, the techniques predicted the likelihood and severity of an automobile collision using road geometry, traffic volume, weather, time of day, and previous crash data. Although the two methods are rather easy to understand, they are not the best for real-time monitoring since they simply display summary data and are unable to record motion conflict at any given time. Vehicle speed, acceleration, traffic density, trajectory sequence, and contextual data have all recently been used by sensor-based models to forecast crash risk or severity [16]. It has been demonstrated that accident risk models require precise multisource perception before prediction, and cooperative raw sensor fusion has been used to give a robust ground-truth generation for intelligent driving settings [17]. Research on sensor-independent fusion navigation has also demonstrated that, instead of employing preset weights, robust fusion should modify sensor contributions based on their reliability [18]. Research on camerasProbabilistic fusion can lower uncertainty in road-scene interpretation, as demonstrated by LiDAR fusion [19]. The multi-view sensor's alignment suggests that traffic data from dispersed devices needs to be spatially calibrated [20]. Although many-resolution sensor fusion structures are often optimized for detection rather than risk assessment, they improve object perception in complex road circumstances [21]. Research on autonomous driving decision-making has also employed cross-semantic sensor fusion to acquire behavior-level knowledge through modality interaction [22]. However, a lot of studies on traffic accidents still rely on coarse spatial grids or historical traffic data, which makes them less sensitive to short-term conflict situations like abrupt braking, forced merging, or unusual lane changes. A more detailed form that can link specific vehicle behavior at a given moment to more general traffic instability is needed for real-time roadside sensing.

Multimodal Fusion and Spatiotemporal Learning in Intelligent Transportation

Multi-modal data fusion is currently a key component of many technologies needed for autonomous driving and intelligent transportation systems. The traffic situation is observed using camera, radar, LiDAR, GPS trajectory, connected-vehicle messaging, and fixed detector data. A comprehensive framework for multimodal traffic prediction has been put forward, and the model of road-network conditions is enhanced using a variety of data [23]. Congestion prediction has also made use of multi-modal fusion and representation mapping, which are intimately linked to unstable driving behavior and accident exposure [24]. Additionally, multimodal data fusion has been used to predict traffic jams; traffic flow, speed, and contextual signals are all complementary [25]. The issues of significant temporal and spatial dependence in road conditions for traffic prediction are frequently addressed by graph neural networks [26]. Using an attention mechanism, attention-based graph prediction models concentrate on significant nodes and time intervals [27]. The evolution of traffic flow is altered by incident information, which should be explicitly taken into account, as demonstrated by dynamic multi-graph models for traffic accidents [28]. Topology modeling is efficient but computationally demanding in real-time deployment, according to surveys of graph neural networks for traffic prediction [29]. As a result, accident risk assessment differs from general traffic prediction in that it focuses on a rare and safety-critical state. The model should be consistent even if the quality of the sensors changes quickly, and it should be able to differentiate between an abnormal conflict and typical congestion. Cross-modal synchronization, adaptive reliability weighting, spatiotemporal interaction modeling, and computationally efficient risk assessment are all essential components of a workable system [30].

Multimodal Real-Time Risk Assessment Framework

Multimodal Sensor Data Representation

The framework synchronously gathers information from environmental sensors, trajectory probes, millimeter-wave radars, traffic-flow detectors, and roadside cameras. They are all distinct kinds of transportation hazards. Lane occupancy, vehicle type, relative location, conflict region, etc. are all provided by visual features. The three characteristics of radar are relative acceleration, radial velocity, and range. Lane-changing behavior and motion continuity are characteristics of a trajectory. Density, average speed, speed variance, and queue length are characteristics of traffic flow. The amount of rain, visibility, light, and road conditions are examples of environmental factors. Initially, the raw observations are sampled every 0.2 seconds and mapped to a common time window of 2.0 seconds. In order to capture abrupt deceleration and local speed changes without going over the inference delay limit for real-time application, ten-time steps of the risk assessment cycle are used.

The multimodal traffic state at time slice t is defined by the concatenation of modality-specific embeddings after temporal alignment.

$$s_t = \Phi_v(v_t) \oplus \Phi_r(r_t) \oplus \Phi_q(q_t) \oplus \Phi_e(e_t) \qquad \text{Eq.(1)}$$

Here, various sensor inputs are simultaneously converted into compact feature vectors by the optical, radar, trajectory-flow, and environmental maps. Since each modality has been normalized and projected independently, the operator denotes feature joining rather than a straightforward raw data stack. Before fusion, all modalities are reduced to 64-dimensional embeddings to prevent the dominance of high-dimensional visual features. A short-term traffic status matrix is then used to depict the matched multimodal sequence. The multimodal traffic accident risk assessment framework's architecture is depicted in Figure 1, which illustrates how sensor observations are transformed into risk-oriented representations.

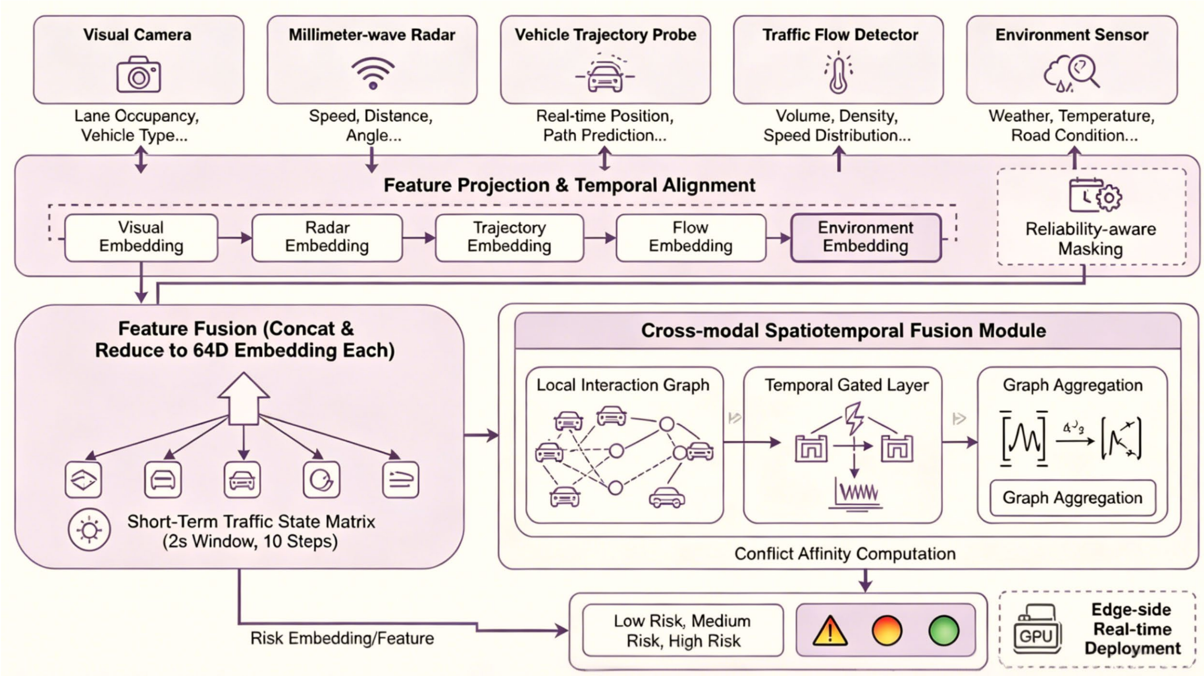

Figure 1. Architecture of the multimodal traffic accident risk assessment framework

The temporal input window is expressed as a sequence of aligned multimodal states.

$$X_T = [s_{t-T+1} s_{t-T+2} \dots s_t] \quad \text{Eq.(2)}$$

T is set to 10 and the total input dimension per time step is 256, as can be seen in the implemented setup. Despite its modest size, this structure retains sufficient data regarding short-term traffic fluctuations. Reliability-aware masking, rather than direct zero-padding, handles missing sensor values; that is, a low-quality camera frame or a missing radar return may be anomalous sensing conditions in and of themselves.

A mode's reliability is determined by its completeness, temporal stability, and compliance with physical rules.

$$\rho_m = \mathrm{sigmoid}(\eta_c c_m + \eta_s u_m + \eta_p p_m) \quad \text{Eq.(3)}$$

The modality is temporally stable and compatible with other observations at that time if the reliability coefficient is comparatively high. For instance, the visual dependability is decreased and the radar branch in fusion is given greater weight if the camera is impacted by glare while the radar velocity is constant. Boost this design's resilience against partial sensor failure, low visibility, and occlusion.

The original traffic variables' physical meaning is likewise preserved in the representation stage. Visual embeddings are limited to maintain lane occupancy, vehicle spacing, and conflict-region information in addition to being employed as abstract image descriptors. The relative velocity and range changes of time-to-collision are preserved by radar embeddings. Acceleration continuity, speed dispersion, and density change are preserved by trajectory and flow embeddings. Because weather and visibility typically serve as risk multipliers rather than direct collision initiators, environmental embeddings are purposefully low-dimensional. As a result, it is easier to understand for further risk analysis and is less likely to overfit the model on a single sensor.

Cross-Modal Spatiotemporal Fusion Module

Both local motion conflicts and global traffic instabilities are addressed by a cross-modal Spatio-temporal Fusion Module. The level of vehicle engagement in a monitored road segment is directly related to the likelihood of accidents. Reduced headway, relatively big speed increases, or erratic lane changes are only dangerous in heavy traffic. As a result, the model initially creates a local interaction graph with vehicles as nodes and edges based on lane connection, distance, and relative velocity. Conflict edges are constructed using camera and radar images, and temporal evolution is provided via trajectory and flow features.

A condensed risk affinity score represents how strongly the two automobiles interact.

https://doi.org/10.64972/jaat.2023v1.287p9e:118-133

$$a_{ij} = \exp(-d_{ij}/\tau_d) \cdot \mathrm{sigmoid}(\Delta u_{ij}/\tau_u) \quad \text{Eq.(4)}$$

If the relative speed is relatively fast and the distance is little, the affinity score will be higher. Update the graph to show the matching model modification at each time step. This local interaction graph is more likely to detect abrupt conflict creation than a fixed road-topology graph, which does not take actual vehicle movement into account.

At every time step, normalized graph fusion is used to collect node-level interaction characteristics.

$$h_i^t = \sigma\left(W_s s_i^t + \sum_j \beta_{ij} W_g s_j^t\right) \quad \text{Eq.(5)}$$

As a result, every vehicle or conflict area will be made aware of any potentially hazardous objects in the vicinity. The surrounding vehicles are also incorporated in the risk representation through interaction weights when a vehicle abruptly cuts in near a small headway distance. Because the intended application demands a quick response, the module is purposefully short, using only one graph aggregation layer and a temporal gated layer rather than a deep graph stack.

A gated update that emphasizes sudden changes in speed, acceleration, and conflict affinity is used to illustrate temporal dependence.

$$g_t = \mathrm{sigmoid}(W_h h_t + W_x s_t + b_g) \quad \text{Eq.(6)}$$

How much of the new conflict information is added to the previous hidden state is decided by the gate. Due to the gate's tiny size in stable traffic, slight sensor noise won't result in risk oscillation. The hidden representation reacts quickly when the gate is opened during fast deceleration or lane changes.

The fused temporal feature is obtained by combining the current conflict-enhanced observation with the historical hidden state.

$$u_t = g_t \odot h_t + (1 - g_t) \odot u_{t-1} \quad \text{Eq.(7)}$$

While avoiding an excessive dependence on a single noisy frame, the above architecture can maintain the model's responsiveness to changes in risk. As a result, during peak hours, the fusion module can maintain a steady risk level and initiate an early alert.

The first is that there are differences between the two sources of accident risk. Aggregation-level graphs typically take into account local spatial relationships and issues such vehicle clusters, lane-changing difficulties, and close following distance. The progression of risk, such as a sharp rise in speed variance or a steady decline in time-to-collision, is the emphasis of the gated temporal layer. The model produces a comparatively high-risk response when the two sources happen at the same time. The risk score will still be relatively high if there is only one cause, such as heavy but steady traffic or an isolated hard braking without any close cars. The two categories are separated in order to lower false alarms in the urban monitoring system.

Real-Time Risk Scoring and Decision Output

Lastly, use the aggregated multi-modal information to create accident risk scores. Low risk, medium risk, and high risk are the three risk output levels. Small distances and lower speeds are indicative of low risk. Unstable local traffic or fresh confrontations are signs of medium danger. Hazardous behaviors include a short time to collision, sudden deceleration, quick lane changes, or notable discrepancy amongst sensor modalities are examples of high-risk behavior patterns. The cross-modal feature fusion and risk scoring method, which incorporates temporal aggregation, reliability weighting, and risk-level decision output, is depicted in Figure 2.

https://doi.org/10.64972/jaat.2023v1.287p9e:118-133

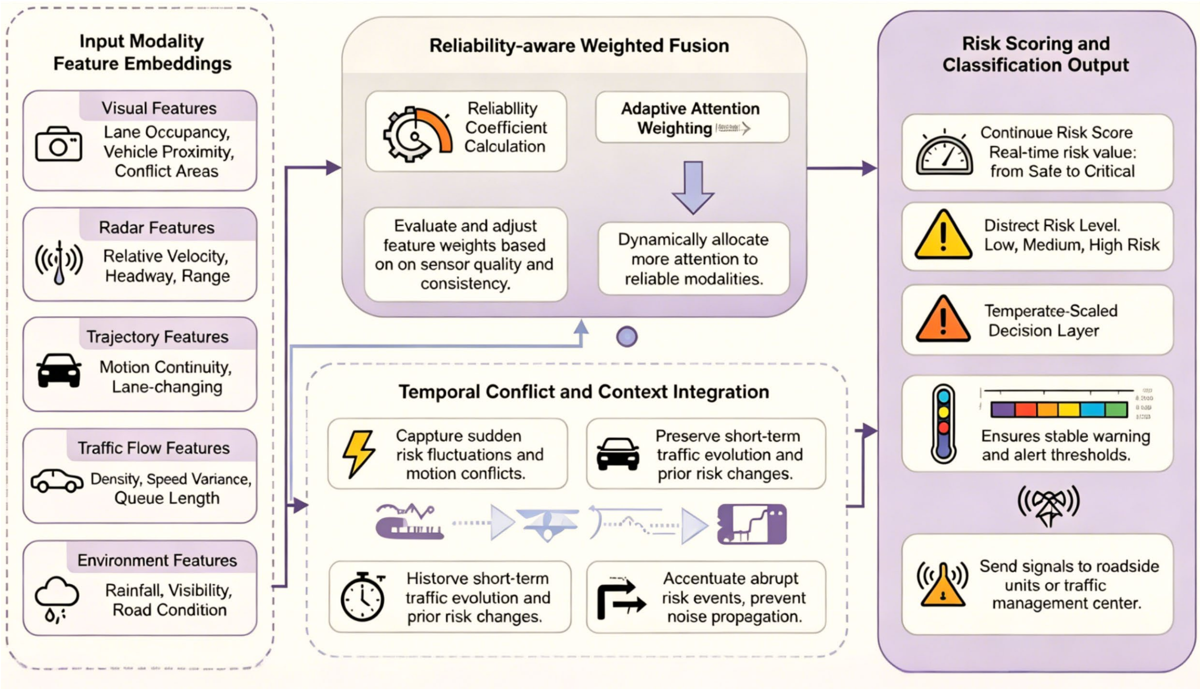

Figure 2. Cross-modal feature fusion and risk scoring mechanism

The modality-weighted fused feature is computed using reliability coefficients and normalized attention weights.

$$f_t = \sum_m \mathrm{softmax}(\rho_m \cdot w_m) \cdot f_m^t \qquad \text{Eq.(8)}$$

By using adaptive weighting, unreliable modalities are not given additional weight. The model will automatically give radar and trajectory features more weight when it is raining and visibility is diminished. Visual and flow elements are employed to make up for the lack of motion information when the radar return is sparse at a junction.

The fused feature and the temporal conflict state are utilized to calculate the real-time risk score.

$$R_t = \mathrm{sigmoid}\left(w_r^T \tanh\left(W_f f_t + W_u u_t\right)\right) \qquad \text{Eq.(9)}$$

The scalar score has a range of [0, 1]. The experiment divides the data into low-, medium-, and high-risk accident categories using thresholds of 0.35 and 0.70. During the validation phase, set the alert's upper and lower boundaries.

A temperature-scaled decision layer determines all of the risk level's classification probability. Temperature scaling can modify the model's probability calibration to improve the output's suitability for traffic control. While a missed high-risk event cannot be disregarded, a calibrated score is necessary to prevent a misleading high-risk alarm that interferes with traffic control.

The training goal is the total of the modality consistency regularization, temporal smoothness constraint, and risk classification loss.

$$L = L_{cls} + \lambda_s \|R_t - R_{t-1}\|_1 + \lambda_c D(f_m; f_n) \qquad \text{Eq.(10)}$$

To make the risk evidence of various modalities consistent, a consistency term is employed, and a smoothness term is added to lessen oscillations in neighboring time steps. AdamW is used for training optimization; early halting is enabled after 80 epochs, the batch size is 64, and the starting learning rate is 0.001. The resulting model is quite efficient, with 2.84 million parameters and an evaluation cycle speed of 31.6 ms on an edge GPU.

Benchmark categorization is not necessary because the output layer is for actual work. A medium-risk output will soon be locally identified and monitored; a high-risk output will trigger warning rules or traffic light coordination; and a low-risk output can be chosen for regular observation. As a result, the model has produced

https://doi.org/10.64972/jaat.2023v1.287p9e:118-133

both a discrete risk category and a continuous risk value. The two outputs—a steady category alert and an uninterrupted value needed by an automatic controller for threshold adjustment—are useful for traffic planners.

Experimental Design and Quantitative Evaluation

Dataset Construction and Implementation Details

The experimental dataset is constructed from multisource traffic monitoring records collected from urban intersections, arterial roads, and expressway merging areas. The dataset contains 186000 synchronized samples, including cameraderived visual descriptors, radar velocity observations, vehicle trajectories, loop-detector traffic flow statistics, and weather sensing variables. Each sample corresponds to a 2.0 s observation window and is labeled as low, medium, or high accident risk according to conflict intensity, time-to-collision, abrupt deceleration, lane-changing instability, and historical incident proximity. The dataset is divided into training, validation, and testing subsets with a $7:1:2$ ratio. Data augmentation includes random sensor dropout, Gaussian radar noise, illumination disturbance, and trajectory jitter. The model is implemented using PyTorch and trained on an NVIDIA RTX 3090 GPU. Edge inference is tested on an embedded GPU platform with 16 GB memory. The normalized risk density for each observation window is calculated as follows.

$$D_T = \frac{1}{T}\sum_{t} \left(\theta_1 C_t + \theta_2 B_t + \theta_3 V_t\right) \quad \text{Eq.(11)}$$

Here, conflict intensity, braking severity, and velocity dispersion are combined to support risk labeling and sample balancing. The final dataset contains 52.6% low-risk samples, 31.8% medium-risk samples, and 15.6% high-risk samples. This distribution reflects practical monitoring conditions where high-risk samples are relatively rare but safetycritical.

The following three criteria are applied for classifying borderline samples in order to lessen the labels' ambiguity. First, if there is a significant speed fluctuation but the vehicle distance is still safe, the sample is not considered high risk. Second, if a sample exhibits both a brief headway and a sharp rise in relative speed at the same time, it is not considered low-risk. Third, unless there is erratic movement or poor sight, bad weather won't immediately raise the risk label. The guidelines can stop the model from learning erroneous associations and align the labels with the real traffic safety reasoning.

Evaluation Metrics for Accuracy and Real-Time Performance

The precision of risk assessment and the speed of real-time operation are the indications. Accuracy, precision, recall, F1-score, AUC, number of parameters, model size, FLOPs, and average inference latency are used to compare the techniques. Single-camera CNN, radar-only classifier, trajectory-only temporal model, early-fusion multilayer perceptron, late-fusion classifier, LSTM fusion network, graph neural network, and the suggested multimodal fusion model are examples of baselines. Because there are fewer samples in the high-risk class than in the low-risk class, a macro-averaged classification score is employed.

$$Acc = \frac{N_{ll} + N_{mm} + N_{hh}}{N_{all}} \quad \text{Eq.(12)}$$

Accuracy measures overall risk-level classification correctness, but it may be biased by the dominant low-risk samples. Therefore, class-balanced recall and F1-score are also reported.

$$F1_{\text{macro}} = \frac{1}{3}\sum_{c} \frac{2P_c R_c}{P_c + R_c} \quad \text{Eq.(13)}$$

Real-time performance is evaluated by the average inference time over 10000 assessment windows. The latency includes feature projection, fusion, graph interaction, temporal update, and risk scoring.

$$T_{avg} = \frac{1}{N}\sum_{i} \left(t_i^{out} - t_i^{in}\right) \quad \text{Eq.(14)}$$

https://doi.org/10.64972/jaat.2023v1.287p9e:118-133

A method is considered suitable for real-time roadside monitoring when its average latency is below 50 ms and its high-risk recall remains above 90%. The proposed model is optimized under this dual constraint rather than maximizing accuracy alone.

In addition to overall metrics, the evaluation records class-level errors because the cost of different errors is unequal. Misclassifying a high-risk state as low risk is more serious than confusing low and medium risk. Therefore, the analysis pays special attention to high-risk recall, medium-risk precision, and the distribution of false alarms. Latency is measured after 100 warm-up iterations to avoid initialization bias, and all inference tests use the same batch size of 1 to match online deployment.

Baseline Models and Experimental Protocol

The same data split, risk labels, and preprocessing procedures are used for training and testing all of the base models. The independent contributions of camera, radar, trajectory, and traffic-flow inputs are evaluated using single-modal baselines. Late fusion averages modality-level prediction scores, while early fusion concatenates all features prior to classification. While the graph neural network baseline solely represents local vehicle interactions, the LSTM fusion baseline takes into account the time-varying features of the data. Reliability-aware modality weighting, dynamic interaction graph aggregation, temporal gated fusion, and calibrated risk scoring are the four that were not part of the prior baselines. Each experiment is conducted five times, and hyperparameters are selected using the validation set. In the results section, a standard deviation and the mean of the reported findings are displayed.

Preventing any benefit for the different input data is the goal of the comparison process. The maximum quantity of data that each baseline's structure can manage is provided. A radar-only model receives a series of ranges and velocities for every object identified, while a camera-only model obtains visual descriptors of every monitored lane. Fusion baselines do not employ dynamic interaction updates or reliability-aware weighting, but they do get the same multimodal input as the suggested model. The configuration will guarantee that the new fusion mode—rather than the addition of additional sensor data—is the cause of the observed performance gain.

Additionally, the experiment's surroundings are not uniform. Turning conflicts, pedestrian interference, and signal-phase changes are all present in intersection samples. Stop-and-go waves and queue spillback are present in arterial-road samples. High-speed lanes and abrupt shifts in relative velocity are characteristics of expressway merging samples. The results analysis will reveal overall performance, modality contribution, scenario discrimination, robustness, and deployment efficiency separately because evaluating all scenarios together would conceal scenario-specific flaws.

Result Interpretation and Comparative Analysis

Overall Risk Assessment Performance

The suggested approach greatly surpasses the single-modal and traditional fusion baseline, with an overall accuracy of 94.1%, a macro F1-score of 92.8%, a macro recall of 93.6%, and an AUC of 96.2% on the test set [31]. Figure 3 compares the overall performance of accident risk assessment; Figure 3(a) displays the accuracy and F1-score of different models; Figure 3(b) displays the precision-recall curve for samples at low, medium, and high risk; and Figure 3(c) displays the confusion matrix of risk-level classification by the suggested model. Although the accuracy of the camera-only baseline is 87.9%, visual occlusion and light changes cause the recall for high-risk samples to drop to 81.4% [32]. With an F1-score of 84.7%, the radar-only baseline performs better in low visibility but is unable to identify trajectory-level instability [33]. Although early fusion has an accuracy of 90.3%, it does not explicitly address the issue of unreliable modalities and keeps redundant features [34]. By using temporal conflict modeling and reliability weighting to differentiate between normal congestion and collision-prone circumstances, the suggested approach reduces the percentage of high-risk misclassification by 28.5% when compared to the early fusion method [35].

The majority of the remaining errors, according to the confusion matrix, are between medium-risk and high-risk states rather than between low-risk and high-risk states. Medium-risk alerts will still be closely monitored, but it is safe enough. The equivalent value for early fusion is 4.9%, while just 1.8% of high-risk samples are incorrectly categorized as low-risk. Thus, it can be said that multimodal temporal evidence is highly effective in detecting a sharp rise in risk.

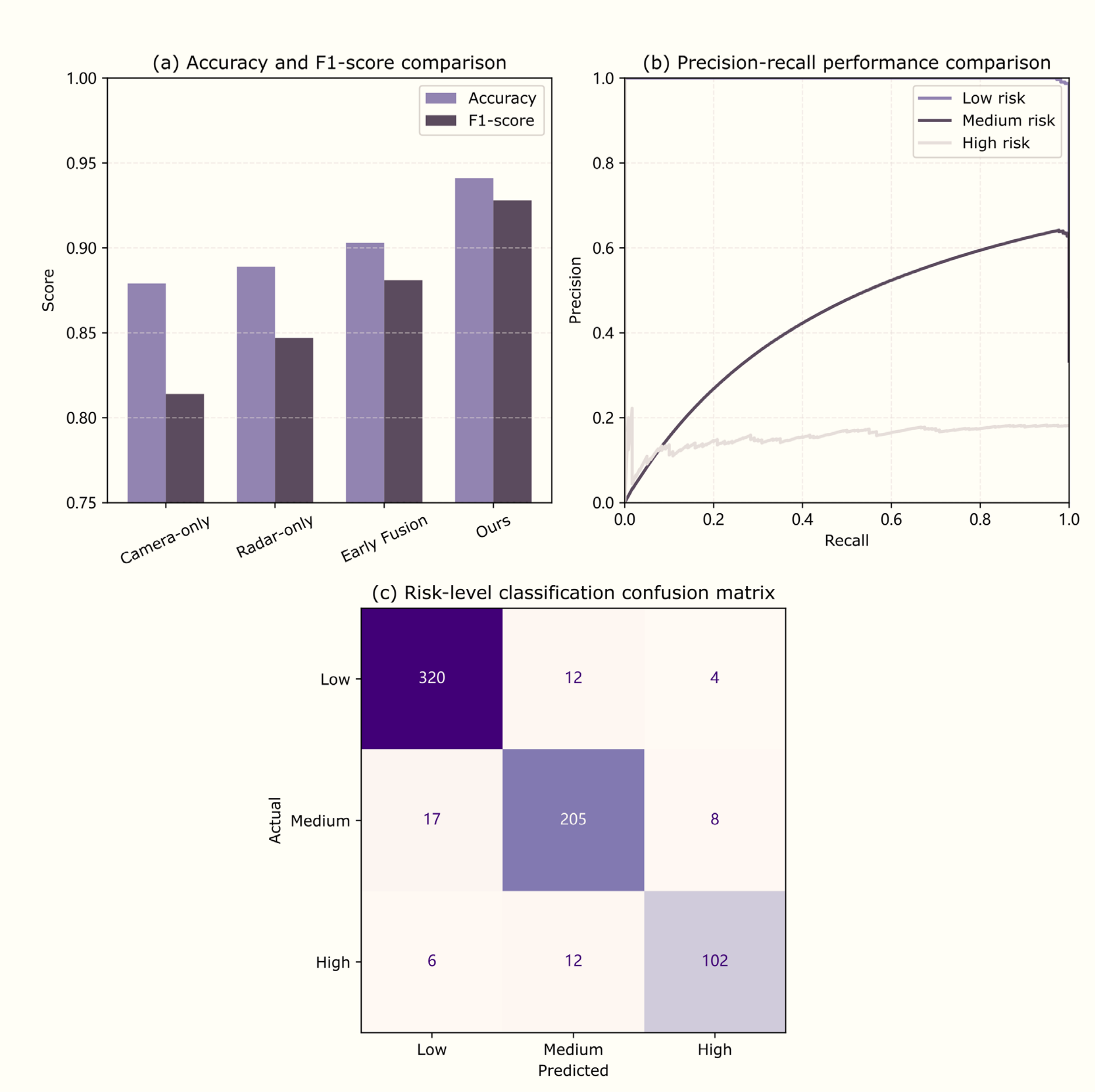

Figure 3. Overall accident risk assessment performance comparison. (a) Accuracy and F1-score comparison. (b) Precision-recall performance comparison. (c) Risk-level classification confusion matrix

Multimodal Fusion Contribution Analysis

According to the contribution study, multimodal fusion works best when there are several interrelated elements contributing to the traffic risk rather than just one aberrant trait [36]. The feature contributions of various sensor modalities are depicted in Figure 4. Figure 4(a) shows that visual features are in charge of lane occupation, vehicle proximity, and visible conflict regions; Figure 4(b) shows that radar and trajectory features account for the recognition of relative velocity, abrupt braking, and motion instability; and Figure 4(c) shows that environmental features are more significant in situations involving rainfall, low visibility, and nighttime illumination. The high-risk recall decreases from 93.8% to 88.1% when radar and trajectory inputs are excluded, and the overall accuracy falls from 94.1% to 91.0% without the visual branch. In favorable weather, environmental factors make for only 4.6% of the average feature weight; in unfavorable weather, this percentage rises to 15.2%. Consequently, the modification suggests that different weights will be assigned to different sources at different periods using adaptive weighting [37].

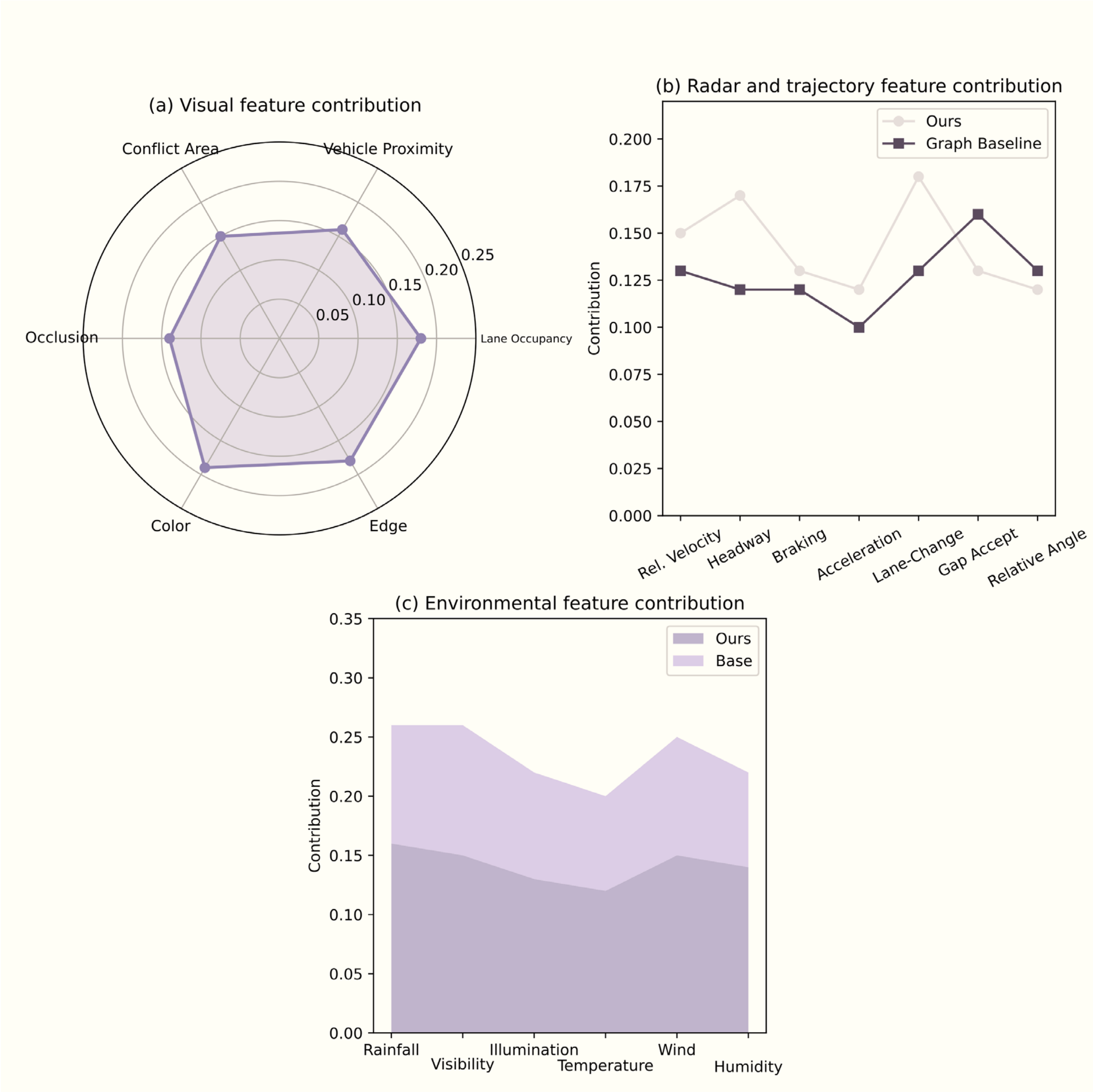

Figure 4. Feature contribution of different sensor modalities. (a) Visual feature contribution. (b) Radar and trajectory feature contribution. (c) Environmental feature contribution

It is also possible to demonstrate how the fusion framework improves risk discrimination using scenario-based analysis [38]. In particular, Figure 5(a) depicts an intersection conflict scenario with frequent turns and crossings, Figure 5(b) depicts a high-speed lane-changing scenario where the headway is short and the relative velocity changes quickly, and Figure 5(c) depicts a low-visibility condition brought on by rain and darkness. By jointly modeling trajectory curvature and visual conflict zone features, the new model has improved the graph baseline for the medium-risk F1 score by 5.7 percentage points at the intersection. Through radar-relative velocity and trajectory continuity, lane-changing situations minimize the number of missed high-risk alarms by 6.4%. By utilizing radar and flow features more extensively, the suggested model still achieves 91.7% accuracy in low visibility, while the camera-only baseline falls to 80.5% accuracy [39].

https://doi.org/10.64972/jaat.2023v1.287p9e:118-133

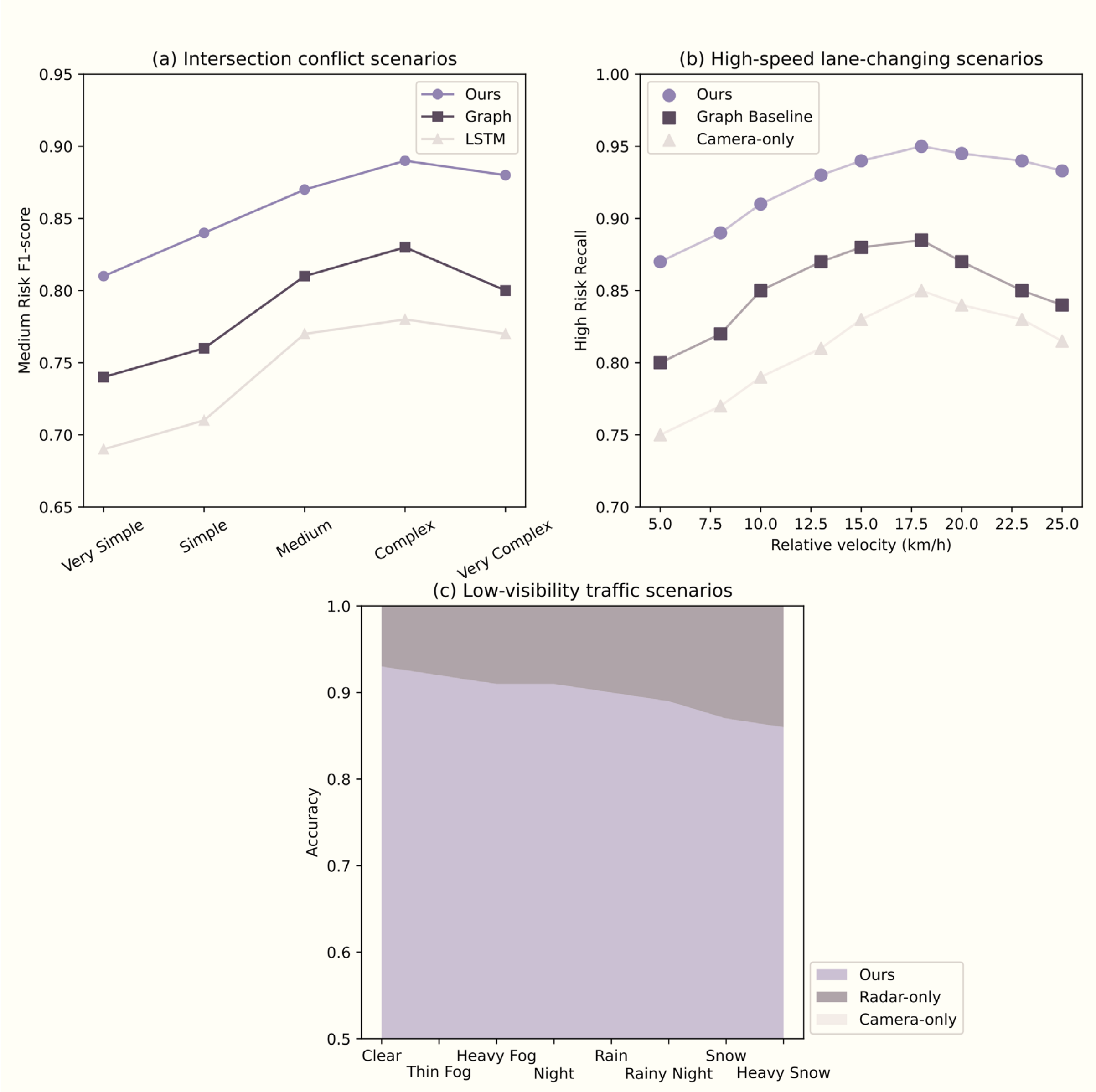

Figure 5. Risk discrimination under complex traffic scenarios. (a) Intersection conflict scenarios. (b) High-speed lane-changing scenarios. (c) Low-visibility traffic scenarios

Robustness and Real-Time Deployment Analysis

In order to assess robustness, one modality is excluded, sensor noise is added, and the sensor sampling rate is reduced [40]. The robustness analysis under sensor noise and missing data is depicted in Figure 6. Specifically, Figure 6(a) shows the performance under visual noise due to blur and illumination disturbance, Figure 6(b) shows the radar disturbance due to velocity jitter and missed returns, and Figure 6(c) shows the condition of missing modalities. The suggested model only loses 2.1 percentage points in F1-score when 20% of the visual frames are disturbed, whereas early fusion loses 6.8 percentage points. The percentage of high-risk recollections decreased from 93.8% to 91.0% under radar disturbance, indicating that the trajectory and visual conflict features can partially offset radar uncertainty. In the absence of the environment, performance decreases by 3.9 percentage points in the rain and is marginally worse in regular weather. The findings demonstrate how the model retrieved sensor information and gave these signals varying weights based on traffic circumstances and signal quality.

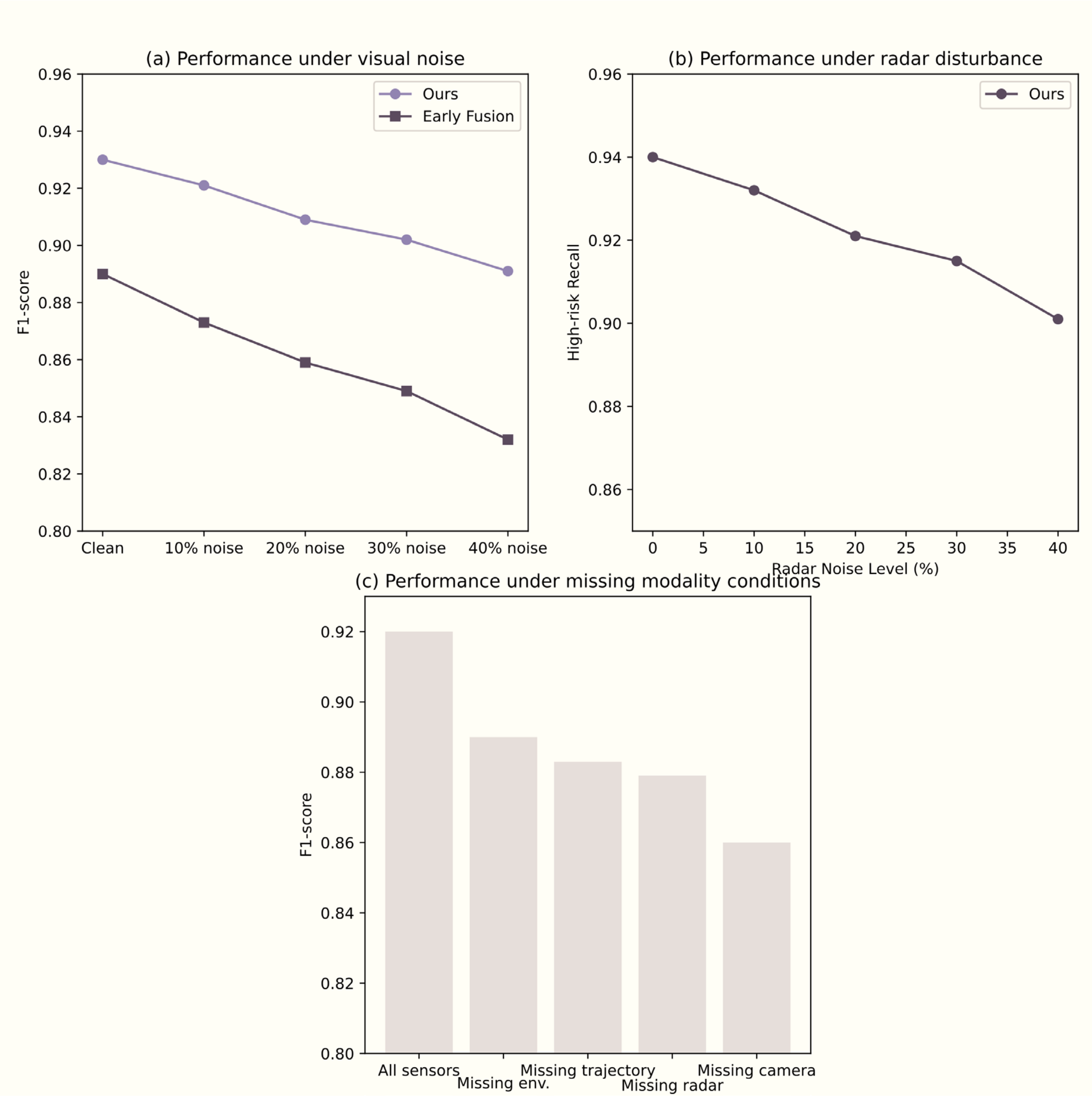

Figure 6. Robustness analysis under sensor noise and missing data. (a) Performance under visual noise. (b) Performance under radar disturbance. (c) Performance under missing modality conditions

The deployment experiment indicates that the suggested model satisfies the real-time monitoring requirements. Figure 7 compares the real-time deployment performance; Figure 7(a) displays the baseline model's inference delay; Figure 7(b) displays the throughput under various sensor sampling rates; and Figure 7(c) displays the edge-device deployment outcomes. The suggested approach has an average delay of 31.6 ms, which is less than that of the deeper graph neural network (57.4 ms) and the LSTM fusion model (48.9 ms). The inbuilt GPU of the model can process 31.2 assessment windows per second, and even after increasing the frame rate from 10 frames per second to 20 frames per second, the throughput remains over 24.5 windows per second. The model is appropriate for the roadside unit with limited processing capacity because it is 11.8MB in size and comprises roughly 2.84 million parameters. A compact fusion design has been employed to minimize memory transfer and enable online risk monitoring for traffic management systems due to the slight latency difference between the desktop and edge GPUs.

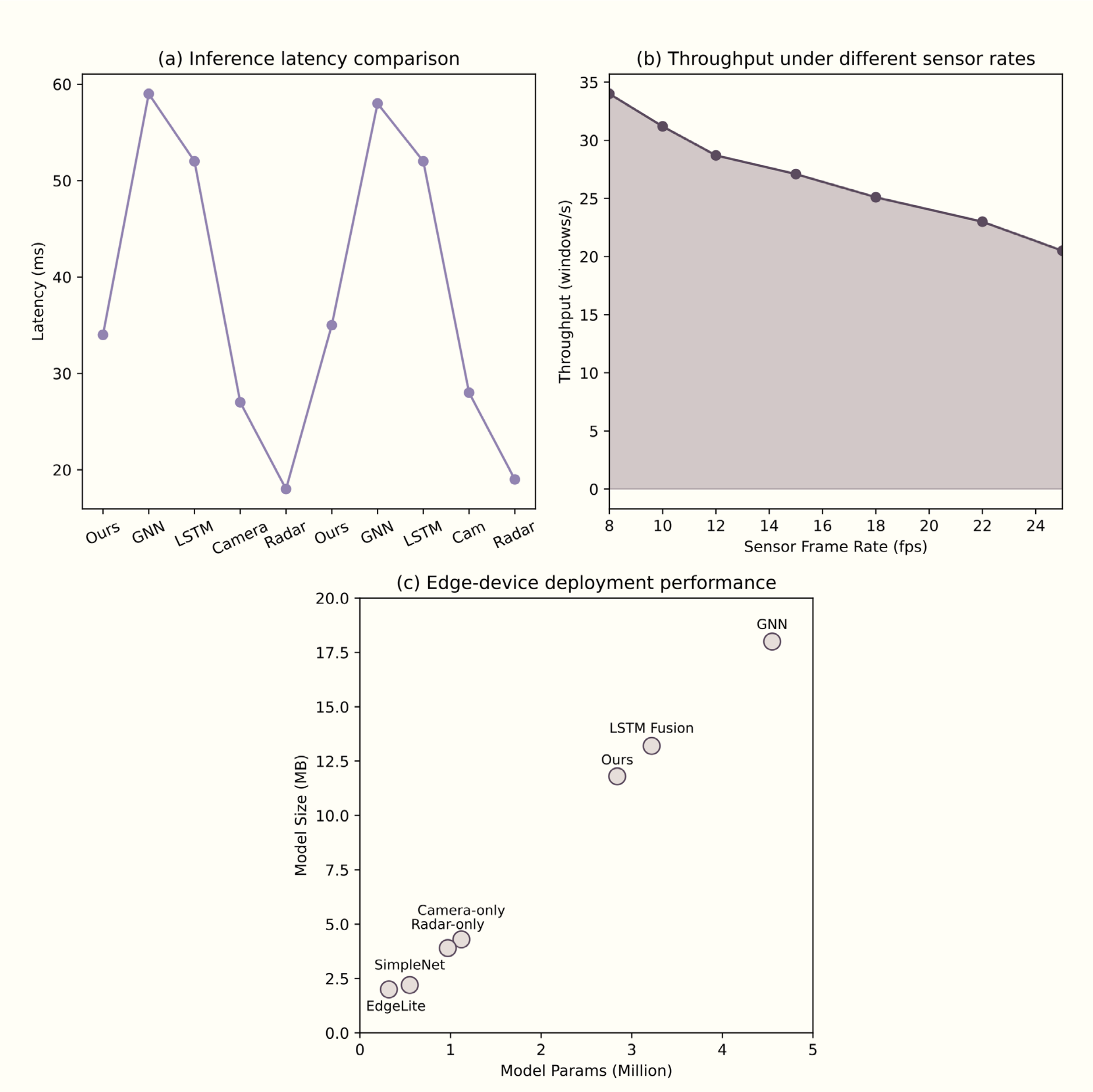

Figure 7. Real-time deployment performance comparison. (a) Inference latency comparison. (b) Throughput under different sensor rates. (c) Edge-device deployment performance

Additionally, latency research reveals that on the embedded platform, the latency for feature projection is 8.4 ms, graph interaction is 9.7 ms, temporal fusion is 6.3 ms, and risk scoring is less than 2.0 ms. The distribution demonstrates that the interaction model, not the final classifier, has the most computing cost. Consequently, sensor-feature caching and sparse conflict graph generation will be the next areas of acceleration. To update the risk state at a frequency appropriate for a standard roadside perception system, the model does not require any optimization.

Because it does not require deep frame-level object reasoning at every stage, the suggested method is more deployment-friendly for online risk monitoring than a heavy video-only network. The tracking results, radar data, and detector statistics that are already accessible from numerous intelligent roadside units can be utilized by the system. Eliminate unnecessary calculations and enhance the warning pipeline's stability at varying sensor rates. The findings also demonstrate that, as a high-accuracy model with high latency might not be feasible for real-time accident prevention, multimodal fusion should be assessed based on both safety performance and operating cost.

Conclusion

A framework for evaluating the risk of traffic accidents in real time using multimodal sensor data fusion. The system as a whole is an integrated spatiotemporal graph of traffic flow statistics, radar motion data, vehicle trajectory information, and camera-based visual perception. According to the findings, when compared to single-modal and conventional fusion baselines, multimodal fusion greatly improves the accuracy of risk-level categorization in high-risk conditions, such as abrupt braking, short distance, lane-changing instability, and bad weather.

The cross-modal spatio-temporal fusion module is the primary cause of the performance enhancement. Dynamic interaction modeling takes into account short-term risk fluctuations and local vehicle conflicts, while reliability-aware modality weighting lessens the effect of noisy or absent sensor data. The new method's average inference speed is quick enough for roadside monitoring to operate in real time, and it has good accuracy and F1 scores. It is appropriate for edge-side traffic safety applications and intelligent transportation systems because it strikes a solid balance between computing economy and discriminatory power.

By including vehicle-to-infrastructure communication, high-definition map restrictions, driver behavior intention, and more comprehensive cross-city traffic data, further work can expand the framework. To enable longer operation in the face of shifting traffic conditions and sensor configurations, model compression, online calibration, and adaptive threshold update should also be investigated.

Author Contributions

Nikodem Baran contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision. Mateusz Robert Cichy contributes to methodology, software, validation, analysis, investigation. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

References

[1] Zhao, H., Zhang, J., Li, X., Wang, Q., & Zhu, H. (2020). Deep learning-based prediction of traffic accident risk in vehicular networks. 2020 IEEE Globecom Workshops (GC Wkshps). https://doi.org/10.1109/gcwkshps50303.2020.9367497

[2] Zhu, L., Li, T., & Du, S. (2019). TA-STAN: A deep spatial-temporal attention learning framework for regional traffic accident risk prediction. 2019 International Joint Conference on Neural Networks (IJCNN). https://doi.org/10.1109/ijcnn.2019.8852212

[3] Rahim, M. A., & Hassan, H. M. (2021). A deep learning-based traffic crash severity prediction framework. Accident Analysis & Prevention, 154, 106090. https://doi.org/10.1016/j.aap.2021.106090

[4] Liyong, W., & Vateekul, P. (2019). Traffic prediction using attentional spatial-temporal deep learning with accident embedding. 2019 4th International Conference on Computational Intelligence and Applications (ICCIA). https://doi.org/10.1109/iccia.2019.00026

[5] Trirat, P., & Lee, J.-G. (2021). DF-TAR: A deep fusion network for citywide traffic accident risk prediction with dangerous driving behavior. Proceedings of the Web Conference 2021. https://doi.org/10.1145/3442381.3450003

[6] Banerjee, K., Bali, V., Sharma, A., Aggarwal, D., & Yadav, A. (2022). Traffic accident risk prediction using machine learning. 2022 International Mobile and Embedded Technology Conference (MECON). https://doi.org/10.1109/mecon53876.2022.9752273

[7] Wu, Y.-W., & Hsu, T.-P. (2021). Mid-term prediction of at-fault crash driver frequency using fusion deep learning with city-level traffic violations. Accident Analysis & Prevention, 150, 105910. https://doi.org/10.1016/j.aap.2020.105910

[8] Zhang, S. (2022). Urban traffic accident prediction research based on meteorological data. 2022 International Conference on Machine Learning and Knowledge Engineering (MLKE). https://doi.org/10.1109/mlke55170.2022.00031

[9] Liyong, W., & Vateekul, P. (2019). Improve traffic prediction using accident embedding on ensemble deep neural networks. 2019 11th International Conference on Knowledge and Smart Technology (KST). https://doi.org/10.1109/kst.2019.8687542

[10] Ma, Z., Mei, G., & Cuomo, S. (2021). An analytic framework using deep learning for prediction of traffic accident injury severity based on contributing factors. Accident Analysis & Prevention, 160, 106322. https://doi.org/10.1016/j.aap.2021.106322

[11] Formosa, N., Quddus, M., Ison, S., & Abdel-Aty, M. (2020). Predicting real-time traffic conflicts using deep learning. Accident Analysis & Prevention, 136, 105429. https://doi.org/10.1016/j.aap.2019.105429

[12] Naseer, A., Nour, M. K., & Alkazemi, B. Y. (2020). Towards deep learning-based traffic accident analysis. 2020 10th Annual Computing and Communication Workshop and Conference (CCWC). https://doi.org/10.1109/ccwc47524.2020.9031235

[13] Zhao, H., Li, X., Cheng, H., & Zhang, J. (2022). Deep learning-based prediction of traffic accidents risk for Internet of vehicles. China Communications. https://doi.org/10.23919/jcc.2022.02.017

[14] Huang, T., Wang, S., & Sharma, A. (2020). Highway crash detection and risk estimation using deep learning. Accident Analysis & Prevention, 135, 105392. https://doi.org/10.1016/j.aap.2019.105392

[15] Chen, X. (2021). Traffic crash severity prediction with deep learning. Journal of Physics: Conference Series, 1883(1), 012141. https://doi.org/10.1088/1742-6596/1883/1/012141

[16] Bao, J., Liu, P., & Ukkusuri, S. V. (2019). A spatiotemporal deep learning approach for citywide short-term crash risk prediction with multi-source data. Accident Analysis & Prevention, 122, 239-254. https://doi.org/10.1016/j.aap.2018.10.015

[17] Li, H., Bai, Q., Zhao, Y., & Qu, Z. (2021). TSDCN: Traffic safety state deep clustering network for real-time traffic crash prediction. IET Intelligent Transport Systems. https://doi.org/10.1049/itr2.12011

[18] Gu, X., Cai, Q., Lee, J., & Xiang, Q. (2020). Proactive crash risk prediction modeling for merging assistance system at interchange merging areas. Traffic Injury Prevention, 21(4), 234-240. https://doi.org/10.1080/15389588.2020.1734581

[19] Adetiloye, T., & Awasthi, A. (2019). Multimodal big data fusion for traffic congestion prediction. In Multimodal Analytics for Next-Generation Big Data Technologies. Springer. https://doi.org/10.1007/978-3-319-97598-6_13

[20] Zhou, B., Liu, J., Cui, S., & Zhao, Y. (2022). Large-scale traffic congestion prediction based on multimodal fusion and representation mapping. 2022 IEEE 9th International Conference on Data Science and Advanced Analytics (DSAA). https://doi.org/10.1109/dsaa54385.2022.10032443

[21] Godhbani, S., Elkosantini, S., Suh, W., & Lee, S. M. (2022). ADL based framework for multimodal data fusion in traffic jam prediction. 2022 14th International Conference on Software, Knowledge, Information Management and Applications (SKIMA). https://doi.org/10.1109/skima57145.2022.10029488

[22] Zhang, A., Liu, Q., & Zhang, T. (2022). Spatial-temporal attention fusion for traffic speed prediction. Soft Computing. https://doi.org/10.1007/s00500-021-06521-7

[23] Fan, S., Li, J., Lv, Z., & Zhao, A. (2021). Multimodal traffic travel time prediction. 2021 International Joint Conference on Neural Networks (IJCNN). https://doi.org/10.1109/ijcnn52387.2021.9533356

[24] Sun, T., Ma, W., Sun, B., & Wang, L. (2021). A decomposition and attention fusion approach for traffic flow forecasting using multimodal deep learning. CICTP 2021. https://doi.org/10.1061/9780784483565.022

[25] Chen, Z., & Zheng, G. (2022). A bidirectional context-aware and multi-scale fusion hybrid network for short-term traffic flow prediction. Promet - Traffic & Transportation, 34(3). https://doi.org/10.7307/ptt.v34i3.3957

[26] Xiao, H., Xiao, J., Deng, X., & Li, L. (2021). Traffic flow prediction based on traffic and meteorological data fusion in non-stationary environments. 2021 International Conference on Electronic Information Technology and Smart Agriculture (ICEITSA). https://doi.org/10.1109/iceitsa54226.2021.00038

[27] Kaygusuz, N., Mendez, O., & Bowden, R. (2021). Multi-camera sensor fusion for visual odometry using deep uncertainty estimation. 2021 IEEE International Intelligent Transportation Systems Conference (ITSC). https://doi.org/10.1109/itsc48978.2021.9565079

[28] Liu, D. (2020). Driver status monitoring and early warning system based on multi-sensor fusion. 2020 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS). https://doi.org/10.1109/icitbs49701.2020.00013

[29] Natan, O., & Miura, J. (2022). Towards compact autonomous driving perception with balanced learning and multi-sensor fusion. IEEE Transactions on Intelligent Transportation Systems. https://doi.org/10.1109/tits.2022.3149370

[30] Cao, X., Chu, N., Liu, D., & Zhang, Z. (2019). Design of multi-sensor information fusion in deceleration early warning system. 2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS). https://doi.org/10.1109/icitbs.2019.00106

[31] Baek, I., Tai, T.-C., Bhat, M. M., & Ellango, K. (2020). CurbScan: Curb detection and tracking using multi-sensor fusion. 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC). https://doi.org/10.1109/itsc45102.2020.9294345

[32] Kuhner, T., & Kummerle, J. (2019). Extrinsic multi sensor calibration under uncertainties. 2019 IEEE Intelligent Transportation Systems Conference (ITSC). https://doi.org/10.1109/itsc.2019.8917319

[33] Liu, H. (2022). Decoupled traffic spatial-temporal graph neural network for traffic flow prediction. 2022 4th International Academic Exchange Conference on Science and Technology Innovation (IAECST). https://doi.org/10.1109/iaecst57965.2022.10062013

[34] Hu, H., Han, K., & Yin, Z. (2022). Bilinear multi-head attention graph neural network for traffic prediction. Proceedings of the 14th International Conference on Agents and Artificial Intelligence. https://doi.org/10.5220/0010763400003116

[35] Liu, Z., & Tan, H. (2021). Traffic prediction with graph neural network: A survey. CICTP 2021. https://doi.org/10.1061/9780784483565.046

[36] Agafonov, A. (2020). Traffic flow prediction using graph convolution neural networks. 2020 10th International Conference on Information Science and Technology (ICIST). https://doi.org/10.1109/icist49303.2020.9201971

[37] Wang, X., Ma, Y., Wang, Y., Jin, W., Wang, X., Tang, J., Jia, C., & Yu, J. (2020). Traffic flow prediction via spatial temporal graph neural network. Proceedings of The Web Conference 2020. https://doi.org/10.1145/3366423.3380186

[38] Cui, Z., Ke, R., Pu, Z., & Ma, X. (2020). Learning traffic as a graph: A gated graph wavelet recurrent neural network for network-scale traffic prediction. Transportation Research Part C: Emerging Technologies, 115, 102620. https://doi.org/10.1016/j.trc.2020.102620

[39] Fukuda, S., Uchida, H., Fujii, H., & Yamada, T. (2020). Short-term prediction of traffic flow under incident conditions using graph convolutional recurrent neural network and traffic simulation. IET Intelligent Transport Systems. https://doi.org/10.1049/iet-its.2019.0778

[40] Han, X. (2019). Traffic incident detection: A deep learning framework. 2019 20th IEEE International Conference on Mobile Data Management (MDM). https://doi.org/10.1109/mdm.2019.00-22