Journal of Science and Transport Technology
Journal homepage: https://jstt.vn/index.php/en

Enhancing multi-sensor classification robustness under noise via adaptive attention fusion

Nguyen Van Tuan[1], Chu Tien Dat[1], Vu Van Quang[1], Bui Hai Dang[1,2], Nguyen Minh Tri[1], Nguyen Duc Son[3], Vu Toan Thang[1,*]

[1]Precision Engineering & Smart Measurement Lab, School of Mechanical Engineering, Hanoi University of Science and Technology, No 1 Dai Co Viet Street, Hanoi 100000, Vietnam

[2]Faculty of Information Technology, University of Transport Technology, No 54 Trieu Khuc Street, Thanh Liet Ward, Hanoi 100000, Vietnam

[3]Intelligent Geosystems and Infrastructure Research Group, University of Transport Technology, No 54 Trieu Khuc Street, Thanh Liet Ward, Hanoi 100000, Vietnam

Article info
Type of article:
Original research paper

DOI:
https://doi.org/10.58845/jstt.utt.2026.en.6.5.364-377

***Corresponding author:**
Email address:
thang.vutoan@hust.edu.vn

Received: 19/04/2026
Received in Revised Form: 11/09/2026
Accepted: 08/10/2026

Abstract: Vehicle classification is an important part of Intelligent Transportation Systems (ITS). While vision-based sensors struggle in harsh weather, non-vision sensors like PIR and LiDAR offer robust additional data. However, conventional fixed-weight fusion strategies often perform poorly when a single sensor becomes noisy. To address this, we propose an Attention-based Adaptive Sensor Fusion (AASF) model that can dynamically adjust fusion weights using a lightweight attention mechanism. The dataset includes 3,963 samples labeled for four main vehicle types: Buses, Cars, Motorbikes, and Trucks. According to experimental results, the AASF model achieved the highest observed accuracy of 96.3%, compared with 95.8% for CNN fusion and 94.4% for fixed-weight fusion. Under clean conditions, its performance was statistically comparable to CNN fusion, while its main advantage was observed under increasing sensor noise. When compared to data from individual sensors, the combined features from the suggested model show a distinct cluster structure and improved separability between vehicle classes through t-SNE visualization. The work specifically shows that the adaptive attention mechanism retains greater robustness against Gaussian noise, enabling the model to automatically modify weights to prioritize sensor channels that are less impacted by noise. This result confirms the potential of sensor aggregation in improving the efficiency of intelligent transportation systems (ITS).

Keywords: Multi-sensor, data fusion, real-time classification, intelligent transportation systems.

1. Introduction

Intelligent Transportation Systems (ITS) play an increasingly important role in improving traffic efficiency, road safety, and transportation management through automated sensing, communication, and data analysis [1]. Among their core functions, vehicle detection and classification provide essential information for traffic-flow monitoring, adaptive signal control, toll collection, road planning, and intelligent transportation

applications [2-4]. Although vision-based systems can provide rich spatial and semantic information, their performance may be affected by illumination changes, occlusion, nighttime operation, and adverse weather, while image processing can also require considerable computational resources [2-4]. These limitations have encouraged the development of non-vision sensing approaches for reliable and low-cost roadside monitoring.

A variety of non-vision sensors have therefore been investigated for vehicle classification, including magnetic, infrared, ultrasonic, radar, and laser-ranging devices [5-11]. Magnetic sensors can detect disturbances in the local magnetic field caused by passing vehicles and have been successfully applied to vehicle classification and behavior recognition [6,7]. PIR sensors are passive, compact, and inexpensive, and can capture thermal variations generated by passing vehicles, whereas ultrasonic sensors can provide complementary distance-related information [8]. Forward-scattering radar can also generate vehicle-dependent signatures for classification [9]. In contrast, laser-ranging and LiDAR sensors provide geometric measurements associated with vehicle shape and dimensions [10,11]. However, each sensing modality has inherent limitations. PIR signals contain limited structural information, while LiDAR or laser-ranging measurements may be degraded by environmental interference. Consequently, combining complementary sensing information is a practical way to improve classification reliability.

In our previous work, synchronized PIR and Laser Rangefinder (LRF) measurements were combined for roadside vehicle classification [12]. The system uses thermal and geometric information to classify buses, cars, motorbikes, and trucks and achieve an average classification accuracy of approximately 94%. This result demonstrates the benefit of multi-sensor information compared with relying on a single sensing modality. However, the previous approach used a predefined fused representation and did not explicitly estimate the reliability or relative contribution of each sensor during inference [12]. Therefore, when one sensing channel becomes noisy or degraded, corrupted information may still be propagated through the fusion process.

More generally, multimodal machine learning has shown that information from heterogeneous sources can be integrated at the data, feature, or decision level to improve predictive performance [13,14]. Similar concepts have been widely applied in intelligent-vehicle perception using combinations of cameras, LiDAR, radar, thermal sensors, and other methods [15-17]. However, a major challenge is that different sensors can be affected differently by environmental conditions. Under adverse weather or sensor-specific disturbances, the quality of one modality may decrease while another remains reliable [16]. Conventional feature concatenation or fixed-weight fusion cannot clearly respond to such changes because the contribution of each modality remains effectively unchanged.

Recent research has increasingly focused on learnable and adaptive sensor-fusion mechanisms. Methods such as PointPainting, CenterFusion, DeepFusion, TransFusion, BEVFusion, and FUTR3D demonstrate that learning cross-modal relationships can improve the utilization of complementary information and enhance robustness to degraded or incomplete sensing inputs [18-23]. In particular, attention-based fusion provides a mechanism for assigning input-dependent importance to different features or sensing streams. Related attention architectures, including Squeeze-and-Excitation Networks, CBAM, Selective Kernel Networks, and ECA-Net, further show that contributing features can be dynamically recalibrated with relatively low additional computational cost [24-27]. These principles are especially attractive for lightweight roadside monitoring, where the fusion model should remain computationally efficient while still being able to respond to sensor degradation.

Despite these developments, most adaptive fusion methods have been designed for autonomous-driving perception systems using high-dimensional camera images, radar measurements, or LiDAR point clouds [15], [17-23]. Comparatively less attention has been given to lightweight adaptive fusion for compact one-dimensional roadside sensor signals such as PIR and laser-ranging measurements. To address this gap, this study proposes an Attention-based Adaptive Sensor Fusion (AASF) model for vehicle classification using synchronized PIR and LiDAR signals. Separate 1D CNN branches are used to extract thermal and geometric feature representations, while a lightweight attention mechanism dynamically estimates the relative contribution of each sensor before classification. The proposed method is evaluated on 3,963 vehicle samples from four classes: Bus, Car, Motorbike, and Truck. In addition to clean-condition classification performance, robustness is evaluated under progressively increasing sensor noise, and the learned attention weights are analyzed to determine whether the model can reduce the influence of a degraded sensor. The proposed framework therefore extends our previous PIR-LRF classification approach [12] from conventional complementary fusion toward reliability-aware adaptive sensor fusion for more robust ITS applications.

2. Methodology

To overcome the limitations of static fusion networks and black-box deep models, we propose an Attention-based Adaptive Sensor Fusion (AASF) model that can assess the dependability of each sensor. By giving adaptive weights to PIR and LiDAR features, the model can prioritize reliable input signals while reducing the influence of noisy or corrupted sensor inputs.

2.1. Adaptive Attention-Based Fusion Framework

The overall architecture of the proposed AASF model is shown in Fig. 1. Vehicle classification using dual sensor data is formulated as a labelled learning problem. The dataset is defined as

$$D=\{(x_{PIR}^{(i)},x_{LiDAR}^{(i)},y^{(i)})\}_{i=1}^{N} \quad (1)$$

where x_{PIR} represents the PIR thermal time-series signal, x_{LiDAR} represents the LiDAR shape measurement, and y is the ground-truth vehicle class label (Bus, Car, Motorbike, Truck).

The goal is to create a mapping function that predicts the vehicle class from the dual-sensor fused information:

$$f(x_{PIR},x_{LiDAR})\rightarrow y \quad (2)$$

Feature extraction for each sensor is performed using a 1D CNN containing two convolutional stages. The first stage consists of a Conv1D layer, batch normalization, ReLU activation, and max-pooling for temporal downsampling. The second stage consists of a Conv1D layer followed by ReLU activation. Temporal downsampling is performed only in the first stage, while the second stage retains the remaining temporal resolution before global average pooling. Global average pooling and a fully connected layer are then used to generate the final feature representation. The architecture of the CNN feature extractor is summarized in Table 1.

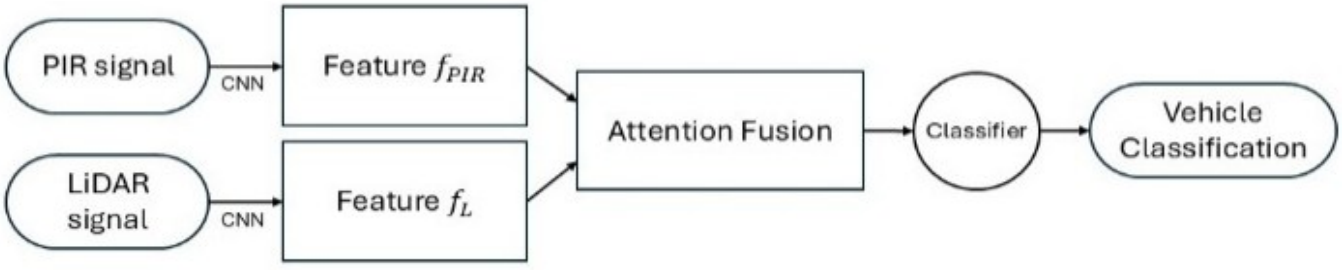


Fig. 1. Architecture of the proposed AASF model

Table 1. 1D CNN feature extractor architecture

Layer	Operation	Kernel Size	Filters	Output
1	Conv1D	3	32	32 × T
2	BatchNorm	–	–	32 × T
3	ReLU	–	–	32 × T
4	MaxPooling	2	–	32 × T/2
5	Conv1D	3	64	64 × T/2
6	ReLU	–	–	64 × T/2
7	Global Avg Pool	–	–	64
8	Fully Connected	–	–	4
9	Softmax	–	–	4

Because PIR and LiDAR input signals capture different physical characteristics—thermal and shape—separate feature extractors are deployed for each sensor. The PIR and LiDAR feature extractors turn the raw signals into latent feature representations:

$$f_{PIR}=\phi_{PIR}(x_{PIR}) \quad (3)$$

$$f_L=\phi_L(x_{LiDAR}) \quad (4)$$

Both extractors output feature vectors of the same dimensionality:

$$f_{PIR}, f_L \in R^d \quad (5)$$

To improve generalization with the limited dataset, several data augmentation techniques are applied, including Gaussian noise injection, temporal shifting, and signal scaling.

Traditional late fusion combines features using fixed weights, such as

$$y=w_1 f_{PIR}+w_2 f_L \quad (6)$$

where w_1 and w_2 denote the fixed weights assigned to the PIR and LiDAR features, respectively. These weights remain constant for all input samples, whereas the proposed AASF model dynamically estimates the contribution of each sensor according to the input features. In the experiments, equal weights were used (w_1=0.6 and w_2=0.4) to provide a neutral static-fusion baseline without preferential weighting of either sensing modality.

However, this static approach cannot adapt when sensor signals become unreliable. To overcome this limitation, we introduce an attention mechanism that dynamically estimates the dependence of each sensor. The process is illustrated in Fig. 2.

First, attention scores are computed by projecting the feature vectors using a learnable weight vector w:

$$e_{PIR}=w^T f_{PIR} \quad (7)$$

$$e_L=w^T f_L \quad (8)$$

Attention scores are normalized using a Softmax function to create adaptive weights:

$$\alpha_{PIR}=\frac{\exp(e_{PIR})}{\exp(e_{PIR})+\exp(e_L)} \quad (9)$$

$$\alpha_L=\frac{\exp(e_L)}{\exp(e_{PIR})+\exp(e_L)} \quad (10)$$

The fusion feature representation is then obtained as



Fig. 2. Attention-based adaptive fusion mechanism

$$f_{fusion}=\alpha_{PIR} f_{PIR}+\alpha_L f_L \quad (11)$$

This allows the model to prioritize reliable sensor inputs while reducing the influence of noisy signals.

The fusion feature vector f_{fusion} is fed into a fully connected layer to predict the vehicle class probabilities:

$$y=\text{softmax}(Wf_{fusion}+b) \quad (12)$$

where W and b denote the classifier parameters.

The entire model—including feature extractors, attention module, and classifier—is trained end-to-end using the cross-entropy loss:

$$L=-\sum y\log(\hat{y}) \quad (13)$$

where y is the ground-truth label, and $\hat{y}$ is the predicted probability distribution.

2.2. Noise Robustness

To evaluate classification performance, we use standard metrics including accuracy, precision, recall, and F1-score, defined as

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN} \quad (14)$$

$$\text{Precision}=\frac{TP}{TP+FP} \quad (15)$$

$$\text{Recall}=\frac{TP}{TP+FN} \quad (16)$$

$$F1=\frac{2\cdot\text{Precision}\cdot\text{Recall}}{\text{Precision}+\text{Recall}} \quad (17)$$

In practical ITS environments, sensor measurements are often affected by environmental noise. Let a sensor measurement be corrupted by additive Gaussian noise:

$$x=x_{true}+\epsilon \quad (18)$$

where $\epsilon\sim N(0,\sigma^2)$. This perturbation leads to the extracted feature representation:

$$f=f_{true}+\delta \quad (19)$$

When processed by the attention mechanism, the attention score becomes

$$e=w^T(f_{true}+\delta)=w^Tf_{true}+w^T\delta \quad (20)$$

The variance of the attention score induced by the noise can be expressed as

$$\text{Var}(e)=w^T\Sigma w \quad (21)$$

where Σ denotes the covariance matrix of the feature perturbation.

During training, the Softmax-based attention mechanism can automatically adjust the sensor's weights according to signal reliability. For example, if the PIR sensor experiences severe interference, the variance of its attention score becomes significantly larger than that of the LiDAR sensor:

$$\text{Var}(e_{PIR})\gg\text{Var}(e_L) \quad (22)$$

Consequently, the attention weights reduce the influence of the noisy sensor:

$$\alpha_{PIR}\downarrow,\alpha_L\uparrow \quad (23)$$

Therefore, the attention-based fusion mechanism can be interpreted as a dynamic weighting process that minimizes the expected classification loss:

$$\min_{\alpha}E[L(y,\alpha_{PIR}f_{PIR}+\alpha_Lf_L)] \quad (24)$$

with

$$\alpha_{PIR}+\alpha_L=1 \quad (25)$$

This formulation explains the improved robustness of the proposed AASF model compared with fixed-weight fusion strategies under noisy conditions.

3. Dataset and Experimental Setup

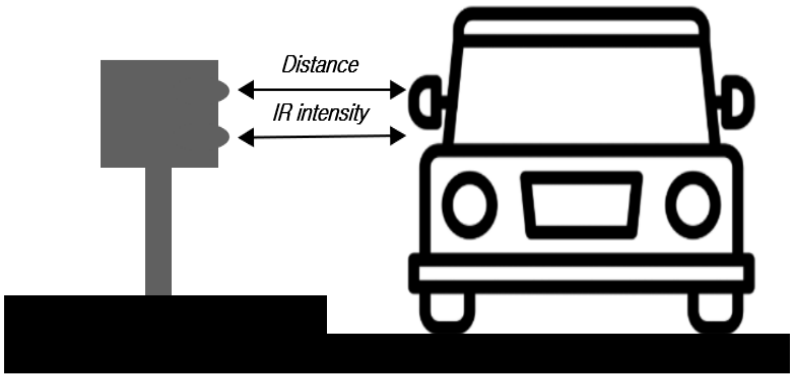


Fig. 3. Roadside PIR–LiDAR sensing setup

To collect the dataset, a sensing system consisting of a Pyroelectric Infrared (PIR) sensor and a LiDAR sensor was deployed beside a roadway in a real traffic environment. The PIR

sensor captures thermal data generated by passing vehicles, while the LiDAR measures vehicle shape. The sensing configuration is illustrated in Fig. 3, and examples of the collected PIR and LiDAR signals are shown in Fig. 4.

The proposed model was implemented using a deep learning framework and trained with the Adam optimizer using an initial learning rate of 0.001. The batch size was 32, and the network was trained for 100 epochs using the categorical cross-entropy loss function. Early stopping based on validation loss was applied to prevent overfitting. All experiments were conducted on a PC equipped with an NVIDIA GTX 1650.

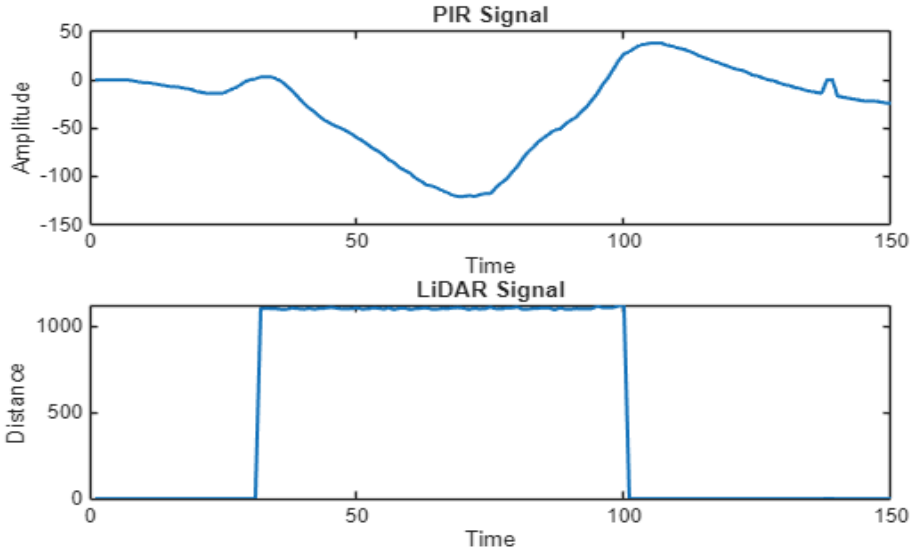


Fig. 4. Example PIR and LiDAR vehicle signals

The dataset contains a total of 3,963 labelled samples, where each sample includes synchronized PIR and LiDAR signals corresponding to a passing vehicle. To establish a reliable baseline before noise evaluation, data collection was conducted on a moderate-density suburban roadway in Hanoi, Vietnam, during daytime hours (08:00–17:00) under standard clear and dry weather conditions. Four vehicle classes are labelled, reflecting realistic traffic proportions: 1,511 Cars (38.1%), 1,005 Motorbikes (25.4%), 932 Trucks (23.5%), and 515 Buses (13.0%). To ensure rigorous model evaluation and prevent overfitting, the dataset was partitioned into a 60-20-20 split utilizing stratified sampling. Specifically, 2,378 samples were used for model training, 793 samples were designated as the validation set for early stopping, and the remaining 792 samples were held out as the testing subset to evaluate final classification performance. Each data sample consists of synchronized PIR and LiDAR measurements corresponding to the same vehicle passage. The two sensing channels provide complementary information: the PIR response reflects the thermal characteristics of the passing vehicle and is also affected by its relative position during detection, whereas the LiDAR signal provides geometric information associated with the same event. Therefore, the two sensor channels are treated as paired measurements rather than as independent sensing systems.

4. Results and discussion

4.1. Performance Evaluation

This section presents the experimental results of the mentioned models. We first compare classification accuracy under clean conditions and then analyze prediction behavior and feature representations. The quantitative comparison is summarized in Table 2.

As shown in Table 2, the single-channel baseline achieves an accuracy of 82.0%. When the complementary PIR and LiDAR information is jointly utilized, the fixed-weight, CNN, and proposed attention-based fusion approaches

achieve accuracies of 94.4%, 95.8%, and 96.3%, respectively. These results indicate that exploiting complementary information from the synchronized PIR and LiDAR measurements improves vehicle classification performance compared with that of the non-fusion single-channel baseline.

Table 2. Accuracy comparison of vehicle classification models

Model	Accuracy
Single-channel baseline (LiDAR and PIR)	82.0%
Attention fusion (Proposed)	96.3%
Fixed weight fusion	94.4%
CNN fusion	95.8%

The obtained results can also be compared with our previous PIR–LRF vehicle classification framework [12]. In the previous study, complementary thermal information from the PIR sensor and geometric information from the laser rangefinder were combined and processed using a lightweight neural network, achieving a reported average classification accuracy of approximately 94.0% [12]. In the present study, the proposed AASF model achieves an overall accuracy of 96.3%. In addition to the higher reported classification accuracy, the main architectural difference is that the proposed framework explicitly estimates sample-dependent sensor contributions through the attention mechanism, whereas the previous model processed the combined sensor representation without adaptive sensor reweighting. Therefore, the present work extends the previous PIR–LRF classification framework from conventional complementary sensor fusion toward reliability-aware adaptive fusion.

A McNemar's test with continuity correction was conducted on the testing predictions to evaluate the performance of the proposed AASF model against the conventional CNN fusion baseline. The analysis yielded χ^2=0.45 and p=0.5023. Since the p-value is greater than 0.05, the 0.5% accuracy difference under ideal noise-free conditions is not statistically significant. This indicates that both models perform comparably well in clean environments; however, the primary advantage of the AASF model emerges under noisy conditions, as will be analyzed in the next section.

Table 3. Comparison with state-of-the-art multi-sensor classification methods

Ref.	Sensor Modality	Classification Method	Reported Accuracy	Adaptive Noise Re-weighting
Odat et al. [8]	PIR + Ultrasonic	Dynamic Bayesian Networks	99.0%	No
Quang et al. [12]	PIR + LRF	Lightweight Neural Network	94.0%	No
Kanona et al. [9]	Forward Scattering Radar	Machine Learning	> 99.0%	No
Proposed (AASF)	PIR + LiDAR	1D CNN + Attention Fusion	96.3%	Yes

While internal baselines confirm the advantage of the adaptive attention mechanism, it is also essential to position the proposed AASF model within the broader context of recent ITS literature. A direct experimental comparison with existing multi-sensor methods is limited by the heterogeneity of sensor modalities and proprietary datasets. For example, methods utilizing PIR and ultrasonic arrays [8], or forward scattering radar [9], process inherently different data structures and physical features.

Table 3 provides a comparative summary of the proposed AASF model against recently published vehicle classification systems on their

respective datasets. While Kanona et al. [9] achieved over 99.0% accuracy using Forward Scattering Radar, such systems generally require specific deployment geometries and higher costs. In the domain of low-cost, non-intrusive sensors, our previous study utilizing PIR and LRF [12] reported an average classification accuracy of approximately 94.0%. The proposed AASF model achieves a highly competitive 96.3% with standard low-cost sensors. More importantly, unlike conventional methods that lack dynamic reweighting, the AASF model's primary contribution lies in its structural capacity to automatically adapt to real-time environmental noise—a critical requirement for continuous outdoor ITS operations.

To further examine prediction behaviour, the confusion matrix of the proposed AASF model is presented in Fig. 5, showing reliable classification across all vehicle categories.

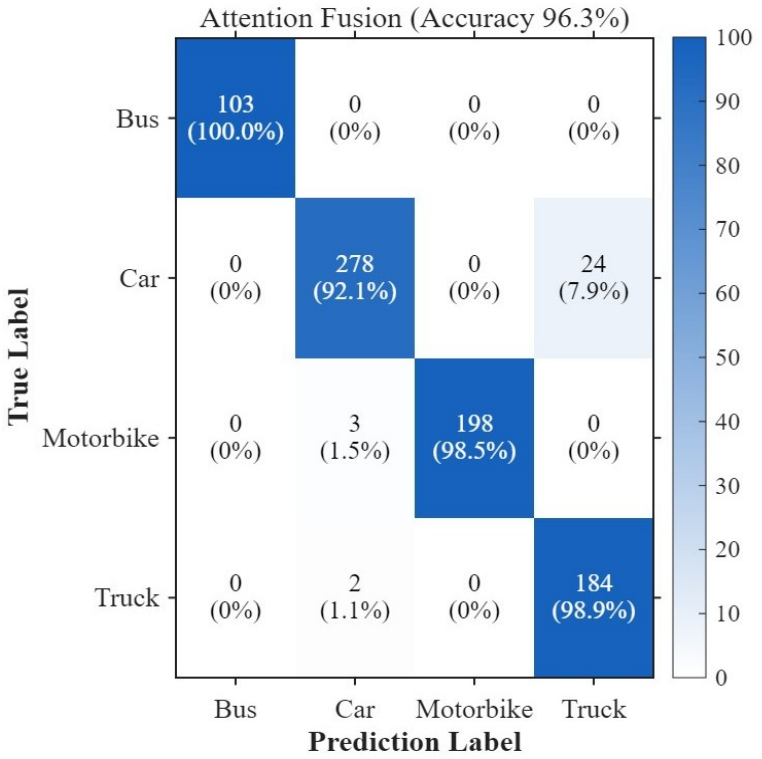


Fig. 5. Confusion matrix of the proposed model

A descriptive class-wise comparison with the previous model [12] provides further insight into the classification behavior. The previous system reported recall values of 100%, 91.96%, 88.9%, and 96.2% for Bus, Car, Motorbike, and Truck, respectively [12]. In the present experiment, the corresponding recall values obtained from Fig. 5 are approximately 100%, 92.1%, 98.5%, and 98.9%. The largest difference is observed for the Motorbike class, for which the recall increases from 88.9% in the previous study to approximately 98.5% in the present experiment. Bus classification remains consistently high, while the Truck class also shows a higher recall in the present results. These observations are consistent with the clearer class separation of the fused feature representation illustrated by the t-SNE visualization in Fig. 6. It should be noted that the two studies employed different model configurations and data partitions; therefore, this comparison is intended to illustrate the progression between the two model designs rather than a strictly controlled head-to-head evaluation.

To analyze feature distinction, t-SNE visualization is used to present the learned feature embeddings in a two-dimensional space. As illustrated in Fig. 6, the PIR features provide moderate class separability, and LiDAR features

exhibit partial overlap between vehicle classes. In contrast, the fusion features demonstrate clearer cluster structures and improved inter-class separation. This visualization suggests that sensor fusion produces a more discriminative feature representation with clearer inter-class separation.

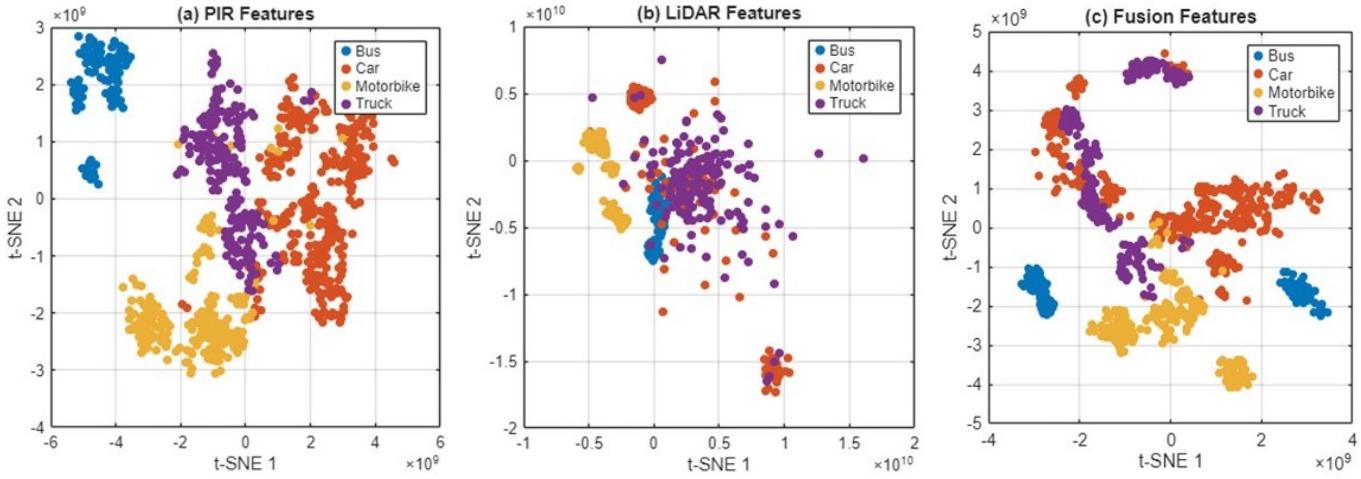


Fig. 6. t-SNE visualization of feature embeddings.

4.2. Noise Robustness Evaluation

While the previous section evaluated performance under clean conditions, practical ITS environments often involve sensor noise and disturbances. Therefore, we evaluate the robustness of the proposed method under increasing levels of Gaussian noise.

To simulate environmental interference, artificial Gaussian noise is added to the sensor signals:

$$x' = x + N(0, \sigma^2) \quad (26)$$

To simulate progressive degradation of one sensing modality, the sensor signals were normalized before noise injection, and zero-mean Gaussian noise was added to the LiDAR channel while the PIR channel was kept unchanged. The Gaussian noise standard deviation was increased as $\sigma = 0, 0.2, 0.4, 0.6, 0.8, 1.0$. The classification accuracy was evaluated on the test set at each noise level. The results in Fig. 7 represent the accuracy obtained for each noise condition rather than repeated-run mean ± SD values.

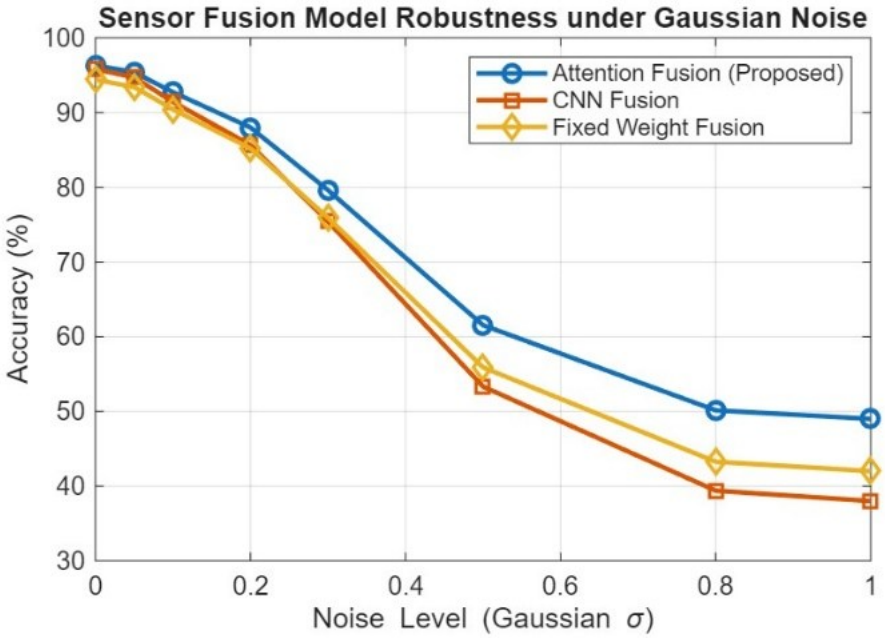


Fig. 7. Accuracy degradation under Gaussian noise

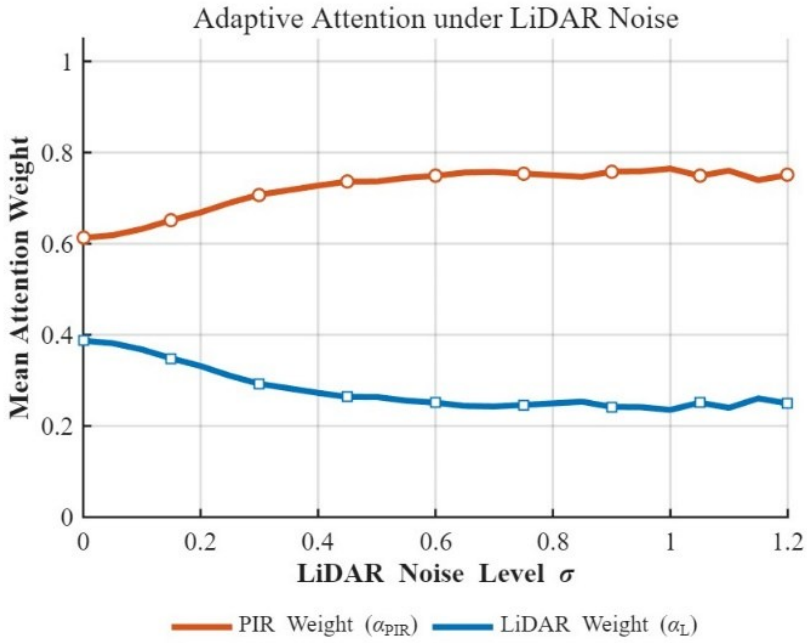


Fig. 8. Attention weights under different noise levels

Table 4. Comparison of fusion strategies

Method	Accuracy	Robustness	Explainability
PIR	low	medium	high
LiDAR	low	medium	high
CNN	high	medium	low
Attention	high	very high	moderate

By gradually increasing the noise variance σ^2, the performance degradation of different models is analyzed, as illustrated in Fig. 7.

As shown in Fig. 7, the attention-based fusion model achieves the highest accuracy under all conditions. In contrast, the CNN fusion model exhibits a faster performance degradation under high noise levels due to the lack of a mechanism to restrain corrupted inputs. Meanwhile, CNN fusion performs slightly better than fixed-weight fusion under clean conditions, but its accuracy degrades more rapidly as the LiDAR noise increases. At higher noise levels, fixed-weight fusion therefore maintains higher accuracy than CNN fusion.

To further analyze this behavior, Fig. 8 shows the average attention weights assigned to the PIR and LiDAR sensors under increasing LiDAR noise levels. As the noise level increases, the model progressively decreases the attention weight of LiDAR while increasing the attention weight of the PIR sensor. Beyond a noise level of around 0.6, the weights become relatively stable, suggesting that the proposed AASF framework can effectively adapt to sensor degradation and maintain robust dual sensor fusion.

The results reveal important trade-offs between classification accuracy, robustness, and interpretability across different fusion strategies. The proposed attention-based fusion achieves the highest accuracy under all conditions. In comparison, the CNN-based fusion model experiences more performance degradation under noisy conditions due to the lack of a mechanism to restrain corrupted sensor inputs. The fixed-weight fusion approach shows a similar degradation trend but remains consistently above the CNN fusion method. By dynamically adjusting attention weights, the proposed fusion model not only improves robustness but also offers partial interpretability of the fusion process. In particular,

the attention coefficients (α_{PIR} and α_L) provide direct insight into the relative contribution of each sensor modality during the fusion process.

The robustness analysis also addresses an important limitation identified in our previous PIR–LRF study [12], where environmental disturbances, including fog and heavy rain, were recognized as factors that could degrade PIR and laser-ranging measurements. In the present study, sensor degradation is examined in a controlled manner by progressively increasing the Gaussian noise applied to the LiDAR channel. As shown in Fig. 7, the advantage of AASF becomes more apparent as the LiDAR signal is increasingly degraded. Meanwhile, Fig. 8 shows that the model progressively reduces the LiDAR contribution and increases its reliance on the unaffected PIR channel. These results indicate that, compared with the previous fixed fusion framework, the main benefit of the proposed AASF model lies not only in its classification performance but also in its ability to adapt the fusion process according to changes in sensor reliability.

The qualitative levels in Table 4 are defined using explicit criteria. For classification accuracy, values below 90% are categorized as low, 90–95% as medium, and above 95% as high. Robustness is assessed from the relative accuracy degradation across the tested LiDAR noise levels in Fig. 7: methods showing intermediate degradation are categorized as medium, whereas the method maintaining the highest accuracy with the smallest degradation across the noise range is categorized as very high. Explainability is assessed based on the accessibility of sensor-specific contributions: direct or predefined sensor contributions are categorized as high, learned attention weights providing partial insight into sensor importance are categorized as moderate, and models without explicit sensor-wise contribution information are categorized as low.

While the simulated Gaussian noise evaluation provides a strong and controlled baseline for verifying adaptive reweighting, it is important to recognize that real-world environmental interference often involves non-Gaussian physical degradation. For instance, adverse weather such as heavy rain or dense fog can cause severe LiDAR beam scattering and multipath reflections, while extreme ambient temperature fluctuations may induce thermal drift in PIR measurements. Priming the model for actual deployment will necessitate evaluating these specific physical noise profiles.

Despite these advantages, several limitations remain. First, the model's capacity for generalization may be constrained by the dataset's limited size when compared to extensive ITS standards. Second, further validation in a variety of settings and traffic situations is necessary because the trials were carried out using a particular roadside sensing arrangement. Future research will concentrate on gathering bigger datasets and investigating more sophisticated fusion architectures.

5. Conclusion

By combining complementary PIR and LiDAR information, this study developed an AASF model to improve vehicle classification performance and robustness under sensor noise. The primary findings from the analysis and experimentation are as follows. First, regarding classification performance, when temperature and shape data are combined, accuracy is far higher than when a single sensor is used alone, reaching 96.3% under typical circumstances. Second, for noise robustness and adaptive weighting, adaptive attention mechanisms are essential for improving the sustainability of the system. In order to maintain steady performance, the model can automatically raise the weight of the PIR channel and decrease the weight of the LiDAR channel as noise levels rise. Finally, regarding feature representation capability, the t-SNE visualization illustrates clearer class separation in the fused feature space, suggesting that the attention-based fusion

produces more discriminative feature representations.

This study has a number of drawbacks, most notably the limited dataset size and the dependence on a single, fixed roadside installation, even though the results are encouraging. Additionally, even while generated noise shows how resilient the model is, it falls short of capturing the intricacy of actual weather dynamics. In future work, our research will focus on expanding the dataset to include diverse traffic scenarios captured across multiple locations during actual adverse weather conditions (e.g., physical LiDAR scattering in rain). Training on these diverse datasets, alongside exploring advanced fusion architectures, will further improve the model's generalization capabilities for real-world ITS deployment.

Acknowledgments

This work was supported by the Hanoi University of Science and Technology under a project (T2024-TD-026). The authors are grateful for the funding.

Conflicts of Interest

The authors declare no conflicts of interest.

Declaration of Generative AI

During the preparation of this work, the authors used Google Gemini in order to improve the readability and language of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

[1] T. Gong, L. Zhu, F.R. Yu, T. Tang. (2023). Edge Intelligence in Intelligent Transportation Systems: A Survey. *IEEE Transactions on Intelligent Transportation Systems*, 24(9), 8919–8944. doi: 10.1109/TITS.2023.3275741

[2]. A. Gholamhosseinian, J. Seitz. (2021). Vehicle Classification in Intelligent Transport Systems: An Overview, Methods and Software Perspective. *IEEE Open Journal of Intelligent Transportation Systems*, 2, 173–194. doi: 10.1109/OJITS.2021.3096756

[3]. M. Won. (2020). Intelligent Traffic Monitoring Systems for Vehicle Classification: A Survey. *IEEE Access*, 8, 73340–73358. doi: 10.1109/ACCESS.2020.2987634

[4]. Z. Wang, J. Zhan, C. Duan, X. Guan, P. Lu, K. Yang. (2023). A Review of Vehicle Detection Techniques for Intelligent Vehicles. *IEEE Transactions on Neural Networks and Learning Systems*, 34(8), 3811–3831. doi: 10.1109/TNNLS.2021.3128968

[5]. R.A. Kerekes et al. (2017). Vehicle Classification and Identification Using Multi-Modal Sensing and Signal Learning. 2017 IEEE 85th Vehicular Technology Conference (VTC Spring 2017), 8108568, pp. 1–5. https://doi.org/10.1109/VTCSpring.2017.8108568

[6]. V. V. Quang, N. V. Tuan, B. H. Dang and V. T. Thang. (2025). Real-time Vehicle Behavior Classification using Single 3-Axis Magnetic Field Sensor and Neural Network. *Journal of Science and Transport Technology*, 5(3), 98-111. doi: 10.58845/jstt.utt.2025.en.5.3.98-111

[7]. P. Sarcevic, S. Pletl, A. Odry. (2022). Real-Time Vehicle Classification System Using a Single Magnetometer. *Sensors*, 22(23), 9299. doi: 10.3390/s22239299

[8]. E. Odat, J.S. Shamma, C. Claudel. (2018). Vehicle Classification and Speed Estimation Using Combined Passive Infrared/Ultrasonic Sensors. *IEEE Transactions on Intelligent Transportation Systems*, 19(5), 1593–1606. doi: 10.1109/TITS.2017.2727224

[9]. M.E.A. Kanona et al. (2022). A Machine Learning Based Vehicle Classification in Forward Scattering Radar. *IEEE Access*, 10, 64688–64700. doi: 10.1109/ACCESS.2022.3183127

[10]. K. Lakshmanan, M. Roach, C. Giannetti, S. Bhoite, D. George, T. Mortensen, M. Manduhu,

B. Heravi, S. Kariyawasam, X. Xie. (2023). A Robust Vehicle Detection Model for LiDAR Sensor Using Simulation Data and Transfer Learning Methods. *AI*, 4(2), 461–481. doi: 10.3390/ai4020025

[11]. J. D. Choi and M. Y. Kim. (2023). A Sensor Fusion System with Thermal Infrared Camera and LiDAR for Autonomous Vehicles and Deep Learning Based Object Detection. *ICT Express*, 9(2), 222–227. doi: 10.1016/j.icte.2021.12.016

[12]. V.V. Quang, N.V. Tuan, N.N. Anh, N.T. Dung, V.T. Thang, B.H. Dang. (2026). Vehicle Classification Using Combined Laser Rangefinder and Pyroelectric Infrared Sensors in a Real Dynamic Environment. *Journal of Science and Transport Technology*, 6(1), 112–123. doi: 10.58845/jstt.utt.2026.en.6.1.112-123

[13]. T. Baltrušaitis, C. Ahuja, L.-P. Morency. (2019). Multimodal Machine Learning: A Survey and Taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2), 423-443. doi: 10.1109/TPAMI.2018.2798607

[14]. J.-H. Choi, J.-S. Lee. (2019). EmbraceNet: A Robust Deep Learning Architecture for Multimodal Classification. *Information Fusion*, 51, 259–270. doi: 10.1016/j.inffus.2019.02.010

[15]. D. Feng, C. Haase-Schütz, L. Rosenbaum, H. Hertlein, C. Gläser, F. Timm, W. Wiesbeck, K. Dietmayer. (2021). Deep Multi-Modal Object Detection and Semantic Segmentation for Autonomous Driving: Datasets, Methods, and Challenges. *IEEE Transactions on Intelligent Transportation Systems*, 22(3), 1341–1360. doi: 10.1109/TITS.2020.2972974

[16]. M. Bijelic, T. Gruber, F. Mannan, F. Kraus, W. Ritter, K. Dietmayer, F. Heide. (2020). Seeing Through Fog Without Seeing Fog: Deep Multimodal Sensor Fusion in Unseen Adverse Weather. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11679–11689*. doi: 10.1109/CVPR42600.2020.01170

[17]. H. Qian, M. Wang, M. Zhu, H. Wang. (2025). A Review of Multi-Sensor Fusion in Autonomous Driving. *Sensors*, 25(19), 6033. doi: 10.3390/s25196033

[18]. S. Vora, A.H. Lang, B. Helou, O. Beijbom. (2020). PointPainting: Sequential Fusion for 3D Object Detection. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4603–4611*. doi: 10.1109/CVPR42600.2020.00466

[19]. R. Nabati, H. Qi. (2021). CenterFusion: Center-Based Radar and Camera Fusion for 3D Object Detection. *2021 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1527–1536*. doi: 10.1109/WACV48630.2021.00157

[20]. Y. Li et al. (2022). DeepFusion: LiDAR-Camera Deep Fusion for Multi-Modal 3D Object Detection. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1526–1535*. doi: 10.1109/CVPR52688.2022.01667

[21]. X. Bai, Z. Hu, X. Zhu, Q. Huang, Y. Chen, H. Fu and C.-L. Tai. (2022). TransFusion: Robust LiDAR-Camera Fusion for 3D Object Detection with Transformers. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1080–1089*. doi: 10.1109/CVPR52688.2022.00116

[22]. T. Liang, H. Xie, K. Yu, Z. Xia, Z. Lin, Y. Wang, T. Tang, B. Wang and Z. Tang. (2022). BEVFusion: A Simple and Robust LiDAR-Camera Fusion Framework. *Advances in Neural Information Processing Systems 35 (NeurIPS 2022). 10421-10434*. doi: 10.52202/068431-0757

[23]. X. Chen, T. Zhang, Y. Wang, Y. Wang, H. Zhao. (2023). FUTR3D: A Unified Sensor Fusion Framework for 3D Detection. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 172–181*. doi: 10.1109/CVPRW59228.2023.00022

[24]. J. Hu, L. Shen, G. Sun. (2018). Squeeze-and-

Excitation Networks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7132–7141*. doi: 10.1109/CVPR.2018.00745.

[25]. S. Woo, J. Park, J.-Y. Lee, I.S. Kweon. (2018). CBAM: Convolutional Block Attention Module. *European Conference on Computer Vision (ECCV), pp. 3–19*. doi: 10.1007/978-3-030-01234-2_1

[26]. X. Li, W. Wang, X. Hu, J. Yang. (2019). Selective Kernel Networks. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 510–519*. doi: 10.1109/CVPR.2019.00060

[27]. Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, Q. Hu. (2020). ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11531–11539*. doi: 10.1109/CVPR42600.2020.01155