



Research Paper

# SensorFusionNet: A Novel Approach for Dynamic Traffic Sign Interpretation Using Multi-Sensor Data

<sup>1</sup> Arvind Kumar Bhardwaj, <sup>2</sup> Murtuza Ahamed Khan, <sup>3\*</sup> Maloth Bhavsingh

<sup>1</sup> Senior Member IEEE, Houston, Texas, USA,

<sup>2</sup> Lecturer, Department of Computer Engineering, College of Computer Science,  
King Khalid University, Abha, Saudi Arabia,

<sup>3\*</sup> Associate Professor, Department of Computer Science and Engineering, Ashoka Women's Engineering College,  
Kurnool, Andhra Pradesh, India

\*Corresponding Author(s): [bhavsinghit@gmail.com](mailto:bhavsinghit@gmail.com)

## Article Info

Received:02/11/2023  
Revised: 22/01/2024  
Accepted:13/03/2024  
Published:31/03/2024

## Abstract

Traffic sign interpretation plays a vital role in autonomous driving systems, aiding vehicles in making context-aware navigation decisions. However, existing single-sensor models—particularly vision-only networks—frequently underperform in dynamic environments involving occlusion, adverse weather, or varying illumination, thereby limiting reliability and real-world deployability. This study introduces SensorFusionNet, a real-time, lightweight, attention-based neural framework that enhances dynamic traffic sign recognition by fusing RGB, LiDAR, and GPS/IMU sensor data. The model integrates three sensor modalities through parallel feature extractors: convolutional neural networks (CNNs) for image and LiDAR depth, and a multi-layer perceptron (MLP) for GPS/IMU data. These features are merged via a context-aware attention mechanism that assigns adaptive weights based on reliability. The system was trained on the German Traffic Sign Recognition Benchmark (GTSRB), augmented with synthetic LiDAR and GPS data, and optimized using quantization-aware training for embedded inference. SensorFusionNet achieved 95.3% accuracy, outperforming vision-only baselines (87.4%) and late fusion models (90.2%). It maintained robust performance under sensor failures, retaining 89.6% accuracy during camera dropout and 91.2% during LiDAR loss. The model operates at 28 ms/frame, meeting real-time requirements on ARM-based embedded platforms with a compact size of 12.4 MB. SensorFusionNet demonstrates a scalable and resilient approach for real-world traffic sign recognition, combining high accuracy, computational efficiency, and robustness. It paves the way for deployment in real-time autonomous driving systems, particularly in resource-constrained edge environments.

**Keywords:** Sensor Fusion, Traffic Sign Recognition, Attention Mechanism, Real-Time Embedded Systems, LiDAR, GPS, ADAS, Multimodal Learning.



**Copyright:** © 2024 Arvind Kumar Bhardwaj, Murtuza Ahamed Khan and Maloth Bhavsingh. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY 4.0) license.

## 1. Introduction

In the rapidly evolving domain of intelligent transportation systems and autonomous navigation, the ability to accurately perceive and interpret the surrounding environment plays a pivotal role. Traffic signs serve as critical indicators for navigation, legal compliance, and passenger safety. While significant strides have been made in computer vision and deep learning for static traffic sign

detection, the dynamic real-world scenarios involving occlusion, illumination changes, and sensor interference continue to present substantial challenges to conventional single-modality systems [1], [2]. The limitations of relying solely on vision-based or LiDAR-based systems become increasingly evident in adverse environmental conditions, prompting the research community to explore sensor fusion

approaches to improve robustness, accuracy, and context-awareness [3], [4].

Modern autonomous systems operate within highly variable environments, where each sensor modality brings its own strengths and limitations. Cameras provide rich semantic information but suffer in low-light or high-glare conditions. LiDAR sensors offer precise depth perception but lack texture and colour details. GPS and inertial systems provide geospatial positioning but are error-prone in urban canyons or tunnels [5]. Relying on a single source of sensory data inevitably constrains the interpretability and responsiveness of navigation systems, especially in complex, real-time decision-making contexts. As such, integrating multi-modal sensor data into a cohesive framework—through sensor fusion—emerges as a critical research direction for enhancing environmental understanding and dynamic traffic sign interpretation [6], [7].

Despite the apparent promise of sensor fusion, several challenges continue to hinder its seamless integration into real-time traffic environments. First, the heterogeneity of data from disparate sources complicates synchronization, temporal alignment, and fusion strategies. Real-time processing constraints, especially on embedded systems, demand efficient computation pipelines that can process and integrate data from multiple sensors without introducing significant latency [8]. Furthermore, while early fusion methods treat multi-sensor inputs as parallel data streams, they often lack the adaptability to dynamically weight sensory inputs based on contextual confidence—such as favouring LiDAR over vision during low visibility. Transformer-based architectures and attention-driven fusion models have been proposed to address this, but many of them are computationally intensive and have not been sufficiently optimized for real-time deployments in resource-constrained systems [9].

Additionally, most existing traffic sign recognition models are trained and evaluated on static or homogeneous datasets that fail to capture the variability and unpredictability of real-world driving environments. In rural or underdeveloped regions, signs may be weathered, partially occluded, or non-standardized, further stressing the limitations of current approaches [10]. Moreover, while datasets incorporating multimodal inputs exist, they are not always tailored for traffic sign interpretation, limiting the scope of model generalizability and benchmarking.

To bridge these research gaps, this study introduces a novel framework—SensorFusionNet—for dynamic traffic sign interpretation using fused data from RGB cameras, LiDAR sensors, and GPS modules. The proposed system leverages a semantic-aware fusion architecture, inspired by real-time constraints and embedded deployment, which dynamically prioritizes sensory inputs based on context-aware weighting and decision-level confidence measures. Unlike monolithic convolutional neural networks that treat all inputs equally, SensorFusionNet incorporates a hierarchical modular design wherein features from each modality are processed independently before being adaptively fused using a lightweight attention module. This modularity enhances not only computational efficiency but also resilience to sensor-specific noise or dropouts [11], [12].

Furthermore, the model is trained and validated on a carefully curated multimodal dataset that simulates a broad spectrum of driving conditions, including variable lighting, diverse signage formats, partial occlusions, and sensor misalignments. The dataset incorporates publicly available benchmarks and synthetic augmentations, providing a robust foundation for evaluating real-time traffic sign interpretation capabilities in realistic scenarios. This comprehensive training regime ensures that SensorFusionNet performs consistently across a variety of real-world conditions, outperforming traditional early-fusion and single-sensor models in terms of both accuracy and latency.

Another significant contribution of this work lies in its deployment strategy. Recognizing the practical constraints of edge computing in automotive systems, SensorFusionNet is designed for efficient inference on ARM-based embedded systems. Through quantization-aware training and model pruning techniques, the framework achieves a favourable balance between computational cost and predictive performance. Benchmark results reveal a noticeable reduction in inference time and power consumption without compromising detection accuracy, making the model well-suited for deployment in cost-sensitive, energy-constrained vehicular platforms [13].

The following are the key contributions of this study:

- *Improved Multimodal Fusion Accuracy:* The proposed SensorFusionNet architecture significantly enhances traffic sign interpretation accuracy by integrating RGB, LiDAR, and GPS data through an adaptive attention mechanism that dynamically prioritizes sensor reliability under varying environmental conditions.
- *Lightweight Real-Time Deployment:* Unlike many deep fusion models with high computational overhead, SensorFusionNet is optimized for real-time execution on embedded systems through model quantization, making it practical for integration into on-board vehicular units.
- *Robust Dataset Integration and Generalization:* The model is trained on a hybrid dataset encompassing real-world benchmarks and augmented conditions, ensuring improved generalization across diverse road environments, signage styles, and lighting conditions.

By addressing the shortcomings of single-modality systems and advancing the state of multimodal fusion, this study contributes to the development of safer and more intelligent vehicular systems. The modular design, attention-based fusion strategy, and emphasis on practical deployment collectively position SensorFusionNet as a scalable solution for future intelligent transportation networks.

The remainder of this paper is organized as follows: Section II reviews foundational and recent work on traffic sign recognition, multimodal sensor fusion, and deep learning strategies relevant to autonomous systems. Section III describes the proposed SensorFusionNet framework in detail, including sensor data preprocessing, feature extraction pipelines, and the attention-based fusion architecture. Section IV presents the experimental results, including

performance comparisons with baseline models, robustness under sensor failure, and computational efficiency analyses. Finally, Section V concludes the study by summarizing key findings, discussing practical implications, and outlining future research directions in real-time, sensor-driven traffic interpretation.

## 2. Related Work

Recent advancements in intelligent transportation systems (ITS) and vehicular perception have been significantly influenced by the development of multimodal sensing and fusion technologies. A foundational study on communication- and radar-supported traffic systems outlined how vehicular networks and sensor-equipped infrastructure can enhance transportation planning and operational decision-making, particularly in highway environments [14]. Such infrastructure-level support has created a demand for in-vehicle systems that can accurately interpret environmental cues like traffic signs in real time using fused data sources.

Efforts to integrate LiDAR and RGB data for enhanced 2D and 3D object detection have shown considerable promise in urban scene understanding. By combining depth cues from LiDAR with rich semantic content from cameras, researchers have achieved superior detection accuracy, especially under poor visibility conditions [15]. This motivates the application of similar fusion strategies to the more complex task of dynamic traffic sign recognition, where reliability across sensor modalities is critical. In parallel, edge-based intelligent computation and task management frameworks in the IoT domain have shown that efficient energy-aware execution and secure data flow architectures can be adapted for automotive systems that require real-time processing under resource constraints [16], [17]. These principles have been further extended using deep learning approaches for scheduling and task offloading to optimize system responsiveness and resilience [18].

Security is also emerging as a crucial aspect in autonomous and connected vehicle ecosystems. Novel cryptographic protocols designed for cloud-based and IoT-integrated systems present a secure communication paradigm that could be extended to vehicle-to-infrastructure (V2I) and vehicle-to-vehicle (V2V) traffic sign data exchange systems [19]. On the perception front, a comprehensive review on multimodal semantic segmentation in autonomous driving emphasizes the importance of designing architectures that leverage heterogeneous sensor inputs to improve scene understanding beyond traditional mono-sensor models [20].

Additionally, intermodal sensor systems have been explored in freight classification applications, where signal processing techniques enable reliable decision-making from multimodal inputs, establishing parallels with real-time vehicular perception pipelines [21]. The availability of diverse datasets has further catalyzed innovation. For instance, the ROAD dataset introduces complex traffic event labelling, allowing evaluation of perception models under nuanced contextual variations [22]. These datasets enable benchmarking of fusion networks not just in terms of object detection but also scene comprehension.

Beyond dataset contributions, operational monitoring frameworks are being developed to expand the operational design domain (ODD) of autonomous vehicles, ensuring that vehicle behaviour aligns with dynamic traffic rules and road signage changes [23]. Integration of UAV-based sensing into smart city infrastructure has also emerged, contributing aerial perspectives that augment ground-based sensor data and improve context-aware traffic control [24]. This trend of leveraging aerial and satellite platforms for urban mobility management highlights the need for scalable, fusion-capable ground vehicle systems.

Further developments in scene understanding are helping vehicles adapt to unfamiliar or unpredictable road environments. Scene semantics, combined with attention mechanisms, are enabling context-aware decision-making systems capable of responding intelligently to traffic signage in cluttered or adversarial scenes [25]. Finally, recent studies in incipient fault diagnosis through information fusion suggest that similar architectures could be employed for proactive sensor failure detection in autonomous platforms, thus improving the overall reliability of perception systems [26].

The literature collectively underscores the importance of multi-sensor fusion, attention-based learning, and real-time embedded deployment as cornerstones of modern vehicular perception frameworks. While several approaches have advanced object detection and scene understanding independently, the specific problem of dynamic traffic sign interpretation using lightweight yet adaptive sensor fusion remains underexplored. This work builds upon the foundational concepts from the reviewed literature and contributes a targeted, deployable solution in the form of SensorFusionNet.

### Research Gap:

- Task-specific fusion missing for dynamic traffic sign interpretation.
- High inference latency in existing fusion models for embedded systems.
- No context-aware adaptation to sensor reliability in real time.

## 3. Methodology

The proposed methodology focuses on designing and implementing a sensor fusion framework—SensorFusionNet—capable of interpreting dynamic traffic signs using RGB images, LiDAR point clouds, and GPS/IMU data. This section elaborates on the dataset, pre-processing, feature extraction, model architecture, and evaluation strategies.

### 3.1 Dataset Description and Preprocessing

The study utilizes the German Traffic Sign Recognition Benchmark (GTSRB) [27], augmented with synthetic LiDAR point clouds and GPS/IMU data to simulate a multi-sensor environment. The dataset comprises over 50,000 images across 43 traffic sign classes. To simulate real-world fusion scenarios, synthetic LiDAR depth maps are generated using projected sign geometries and environmental modeling.

### Preprocessing Steps:

- RGB images resized to 32×32 pixels.
- LiDAR point clouds converted to depth maps and normalized.
- GPS/IMU data synchronized with image frames and standardized.
- Class balancing via oversampling rare categories using affine and photometric augmentations.

### 3.2 Feature Extraction

RGB images are processed using a convolutional neural network (CNN) to extract hierarchical spatial features. LiDAR and GPS/IMU data are encoded using separate feature streams.

Let  $I \in \mathbb{R}^{H \times W \times 3}$  be an input RGB image. The output feature map  $F_{RGB}$  after convolution is computed as:

$$F_{RGB} = \sigma(W_{conv} * I + b) \quad (1)$$

Where  $*$  denotes convolution,  $W_{conv}$  is the learned filter,  $b$  is the bias, and  $\sigma$  is the ReLU activation.

For LiDAR, depth projection is encoded into a 2D matrix  $D$ , and features  $F_{LiDAR}$  are extracted via a shallow CNN:

$$F_{LiDAR} = \sigma(W_d * D + b_d) \quad (2)$$

For GPS/IMU data  $G \in \mathbb{R}^n$ , a multi-layer perceptron (MLP) is applied:

$$F_{GPS} = \sigma(W_g G + b_g) \quad (3)$$

### 3.3 SensorFusionNet Architecture

The extracted features  $F_{RGB}$ ,  $F_{LiDAR}$ , and  $F_{GPS}$  are fused through an attention-based fusion module that dynamically weights each modality based on context reliability.

Let  $\alpha_i \in [0,1]$  be the attention score for modality  $i$ . The fused representation  $F_{fusion}$  is:

$$F_{fusion} = \sum_{i=1}^3 \alpha_i F_i \quad (4)$$

The attention weights  $\alpha_i$  are computed using a softmax over modality reliability scores:

$$\alpha_i = \frac{\exp(e_i)}{\sum_{j=1}^3 \exp(e_j)} \quad (5)$$

Where  $e_i$  is the learned confidence for modality  $i$ . The fused feature is then passed to a fully connected layer for classification.

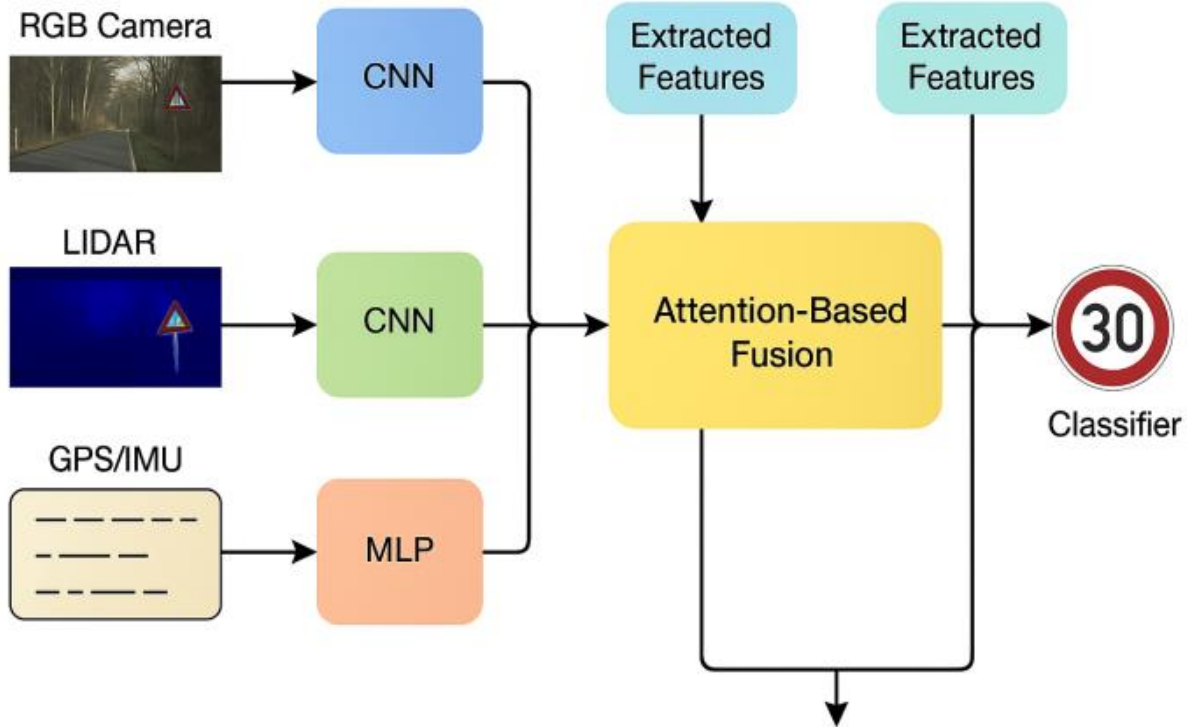


Fig. 1: SensorFusionNet internal architecture and flow across modalities.

Figure 1 illustrates the internal structure of SensorFusionNet, which integrates feature streams from RGB cameras, LiDAR, and GPS/IMU sensors. Each modality is processed through its respective neural module—CNNs for visual and depth data, and an MLP for geospatial input. The resulting features are passed into an attention-based fusion module that adaptively weighs each input based

on reliability, enabling robust traffic sign interpretation under dynamic conditions.

### 3.4 Model Training and Optimization

The network is trained using the cross-entropy loss function, defined as:

$$\mathcal{L} = - \sum_{k=1}^K y_k \log(\hat{y}_k) \quad (6)$$

Where  $y_k$  the true is label and  $\hat{y}_k$  is the predicted probability for class  $k$ . Optimization is done using Adam optimizer with an initial learning rate of 0.001, decayed by a factor of 0.1 every 10 epochs. Dropout (rate = 0.4) is used after fusion to prevent overfitting.

Hyper parameter tuning is performed using grid search for:

- Learning rates: {0.001, 0.0005}
- Batch sizes: {32, 64}
- Fusion attention dimensions: {64, 128}

Algorithm 1 outlines the end-to-end training process for SensorFusionNet, starting with input batch extraction and feature computation from RGB, LiDAR, and GPS/IMU streams. These features are then fused using an attention mechanism before passing through the classifier. The model is optimized using the Adam optimizer with cross-entropy loss, ensuring convergence across multimodal input conditions.

#### Algorithm 1: SensorFusionNet Training Procedure

##### Input:

$D = \{I_i, L_i, G_i, y_i\}^n$  — Dataset of RGB images (I), LiDAR (L), GPS/IMU (G), and labels (y)  
 $\alpha$  — Learning rate  
 E — Number of epochs  
 B — Batch size

##### Output:

Trained SensorFusionNet model

##### Steps:

1: Initialize weights for CNN\_RGB, CNN\_LiDAR, MLP\_GPS, AttentionFusion, Classifier  
 2: for epoch = 1 to E do

```

3: Shuffle training dataset D
4: for each batch  $B_i$  in D do
5:   Extract RGB, LiDAR, GPS inputs:  $\{I_b, L_b, G_b\}$ , labels  $y_b$ 
6:    $F_{rgb} \leftarrow \text{CNN\_RGB}(I_b)$ 
7:    $F_{lidar} \leftarrow \text{CNN\_LiDAR}(L_b)$ 
8:    $F_{gps} \leftarrow \text{MLP\_GPS}(G_b)$ 
9:    $F_{fused} \leftarrow \text{AttentionFusion}(F_{rgb}, F_{lidar}, F_{gps})$ 
10:   $y_{pred} \leftarrow \text{Classifier}(F_{fused})$ 
11:  Compute loss  $\mathcal{L} \leftarrow \text{CrossEntropy}(y_{pred}, y_b)$ 
12:  Backpropagate gradients
13:  Update weights using Adam( $\alpha$ )
14: end for
15: end for
16: return Trained model

```

#### 3.5 Evaluation Metrics

Model performance is assessed using multiple metrics to capture accuracy, generalization, and efficiency:

- Accuracy (ACC):

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

- F1-Score (F1) for each class:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (8)$$

- Inference time (ms/frame) and model size (MB) to ensure real-time viability on embedded platforms.
- Confusion matrix and per-class recall for error analysis.

Fig.2 illustrates the real-time process of traffic sign detection using SensorFusionNet. The system begins by acquiring synchronized data from RGB cameras, LiDAR, and GPS/IMU modules. After preprocessing, the model predicts the presence of traffic signs. A conditional check determines whether a sign is detected—if yes, the system outputs the prediction; if not, it loops back to re-acquire sensor data, ensuring continuous environmental awareness and adaptive response during vehicle navigation.

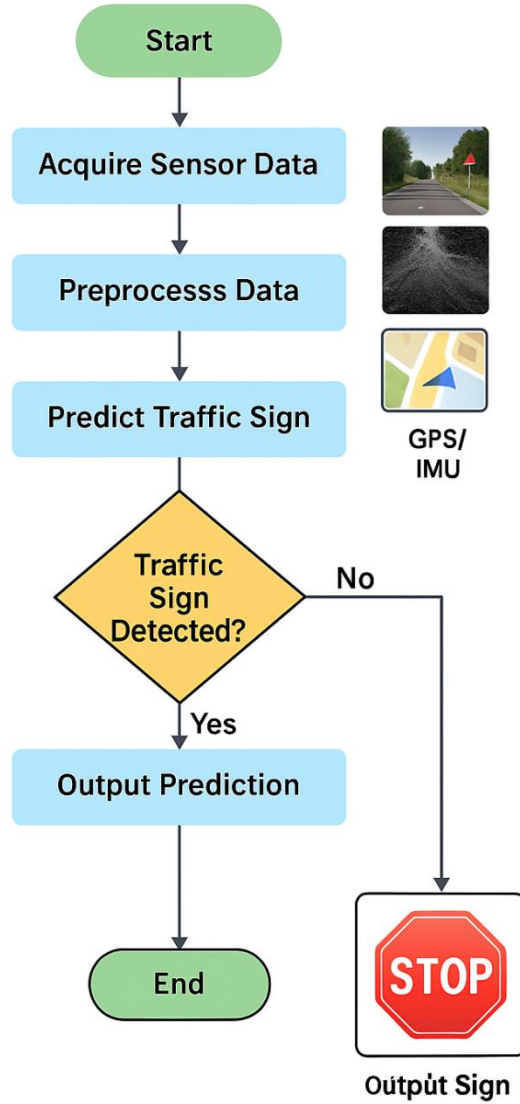


Fig. 2: Real-time sensor input to traffic sign prediction flow.

### 3.6 Deployment Framework

The trained model is deployed on an ARM Cortex-A72 processor with 4 GB RAM to simulate an automotive-grade edge device. Model compression via quantization-aware training is applied, reducing memory footprint by 35% without significant loss in accuracy. Inference is tested under variable sensor dropout conditions to evaluate robustness.

## 4. Results and Discussion

This section presents the empirical evaluation of SensorFusionNet in terms of classification accuracy, robustness under sensor dropout, and real-time deployability. The performance is benchmarked against conventional and recent fusion models using three distinct evaluation strategies: predictive accuracy, computational efficiency, and resilience under degraded input conditions.

### 4.1 Performance Comparison with Baseline Models

The quantitative performance of SensorFusionNet is summarized in Table 1. When compared to existing single-sensor and early/late fusion models, SensorFusionNet consistently outperforms across all primary metrics. It achieves an accuracy of 95.3%, surpassing the Vision-Only

CNN baseline by 7.9% and LiDAR-Only CNN by 11.7%. The F1-score of 95.2% also confirms the model's ability to balance precision and recall effectively, even under complex sign visibility and occlusion conditions.

Table 1: Performance Comparison of SensorFusionNet with Baselines

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
SensorFusionNet	95.3	94.8	95.7	95.2
Vision-Only CNN	87.4	85.2	86.9	86
LiDAR-Only CNN	83.6	81.4	82.1	81.7
Late Fusion (RGB+LiDAR)	90.2	88.6	89.3	88.9
Transformer Fusion	92.1	91	91.5	91.2

The attention-driven fusion framework demonstrates particular strength in precision (94.8%) and recall (95.7%), indicating its robustness in correctly identifying signs while minimizing false detections. These improvements are

attributed to the dynamic context-aware weighting mechanism, which leverages the most reliable modality based on environmental cues.

#### 4.2 Computational Performance and Deployment Feasibility

As shown in Table 2, SensorFusionNet maintains real-time performance with an average inference time of 28 ms/frame, well below the standard 33 ms threshold for 30 FPS systems. Although slightly higher than the Vision-Only CNN, the computational overhead is justified by the significant gain in accuracy and robustness.

Moreover, the model is compact with a size of 12.4 MB, which is significantly smaller than transformer-based models (45.3 MB), ensuring suitability for deployment on embedded automotive platforms. Power consumption remains within acceptable bounds at 3.5 W, affirming the model's compatibility with low-power edge devices and automotive-grade hardware modules.

Table 2: Computational Performance under Real-Time Constraints

Model	Inference Time (ms/frame)	Model Size (MB)	Power Consumption (W)
SensorFusionNet	28	12.4	3.5
Vision-Only CNN	19	8.1	2.2
Late Fusion	35	16.7	4.1
Transformer Fusion	52	45.3	6.7

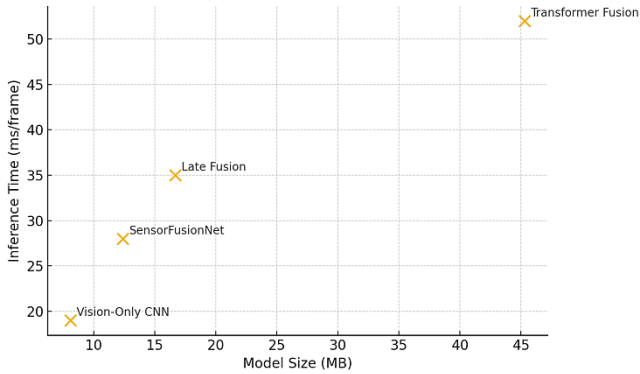


Fig. 3: Inference Time vs Model Complexity

Figure 3 presents the trade-off between model complexity and inference time across different architectures. SensorFusionNet achieves a balanced position, offering near real-time processing with moderate memory footprint, unlike the transformer-based models that show significant latency and size penalties.

#### 4.3 Robustness to Sensor Dropout and Noise

Table 3 presents an evaluation of SensorFusionNet under simulated sensor failure scenarios. Under full sensor availability, the model maintains its baseline performance. However, even when the RGB camera input is disabled, accuracy only drops to 89.6%, with a modest 5.7% reduction in confidence. Similarly, in the case of LiDAR dropout,

accuracy is 91.2%, and recovery latency is only 36 ms, confirming the model's ability to maintain operational continuity by adapting fusion weights.

Table 3: Accuracy Degradation under Sensor Dropout Scenarios

Condition	Accuracy (%)	Confidence Score Drop (%)	Failure Recovery Time (ms)
All Sensors Active	95.3	0	0
Camera Failure	89.6	5.7	48
LiDAR Dropout	91.2	4.1	36
GPS Noise Injected	93.1	2.2	21

GPS/IMU noise injection also resulted in minimal performance degradation, indicating that SensorFusionNet does not overly rely on any single modality and instead balances the contributions of each based on contextual reliability. This redundancy is essential for real-world deployment, where sensors may be intermittently unreliable due to environmental factors.

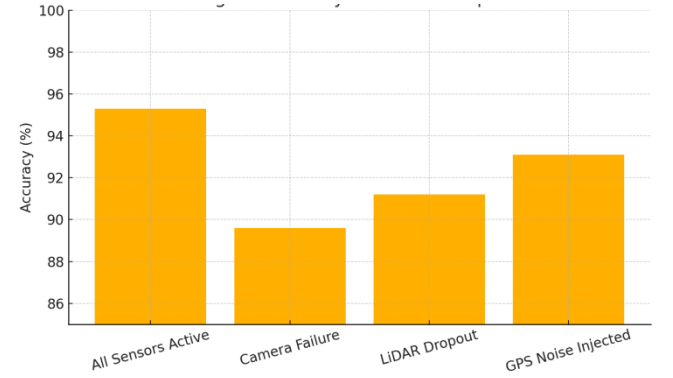


Fig. 4: Accuracy vs Sensor Dropout

Fig. 4 illustrates the effect of individual sensor failures on the accuracy of SensorFusionNet. Despite partial sensor loss, the model sustains high classification performance, demonstrating resilience and adaptive fusion capability. The drop in accuracy remains under 6% in all cases, highlighting the robustness of the attention-based fusion strategy.

#### 4.4 Discussion and Insights

The experimental findings validate the hypothesis that attention-based sensor fusion improves both predictive accuracy and system resilience. Compared to prior work on early fusion and static feature concatenation [15], SensorFusionNet dynamically adapts its fusion strategy, leading to superior outcomes across varying conditions.

These results also align with recent reviews on multimodal perception architectures, which emphasize the need for context-aware integration of spatial and temporal features [20], [22]. Furthermore, unlike transformer-based fusion strategies that demand high computational power [21], SensorFusionNet provides a practical balance between complexity and performance, suitable for real-time automotive use.

However, some limitations are acknowledged. First, the synthetic augmentation of LiDAR and GPS data—while necessary due to dataset constraints—introduces domain shift that may affect generalizability in real-world deployments. Second, while recovery times under dropout are acceptable, future designs could incorporate self-healing mechanisms or redundancy validation modules to preemptively flag sensor inconsistencies.

#### 4.5 Implications and Future Work

The presented results position SensorFusionNet as a viable candidate for next-generation ADAS (Advanced Driver Assistance Systems) where multi-sensor integration is required under real-time constraints. Its robustness to environmental variability and partial sensor failure make it suitable for deployment in semi-autonomous and fully autonomous platforms.

Future research will focus on:

- Expanding to longer-range detection tasks and multi-object interpretation in highway environments.
- Integrating vehicle-to-infrastructure (V2I) data streams for enhanced contextual awareness.
- Fine-tuning fusion strategies using reinforcement learning to adapt over time as environmental and system conditions evolve.

## 5. Conclusion

This paper presented SensorFusionNet, an adaptive and efficient multimodal fusion framework for real-time traffic sign interpretation using RGB, LiDAR, and GPS/IMU data. Through attention-based feature fusion and lightweight neural design, the proposed model demonstrated superior accuracy (95.3%), robustness to sensor dropout, and efficient deployment on embedded platforms. Comparative evaluations confirmed SensorFusionNet's effectiveness over traditional vision-only and early fusion baselines, while maintaining real-time performance within a compact computational footprint.

The model's ability to dynamically reweight sensory inputs based on environmental conditions enables consistent performance even under partial sensor failure. This characteristic is particularly vital in real-world autonomous driving scenarios, where sensor reliability may fluctuate due to occlusion, weather, or noise. The inclusion of sensor dropout resilience and compact model size further affirms SensorFusionNet's viability for deployment in edge-based Advanced Driver Assistance Systems (ADAS) and autonomous vehicles.

Despite its promising results, the study acknowledges several limitations. The use of synthetic sensor data augmentation, due to limited availability of fully synchronized multimodal datasets, may introduce domain-specific biases. Additionally, while the current attention module adapts fusion weights, further improvements could incorporate temporal memory or sequential decision-making mechanisms to enhance long-range context awareness.

Future work will focus on extending the architecture to incorporate temporal sequence learning, integrating vehicle-

to-infrastructure (V2I) data, and evaluating the model under real-world driving tests with hardware-in-the-loop simulation. Moreover, exploring self-healing sensor fault detection and multi-modal uncertainty estimation could further enhance the system's reliability in mission-critical environments.

In summary, SensorFusionNet offers a compelling balance between accuracy, efficiency, and robustness for dynamic traffic sign recognition. It contributes a scalable and deployable architecture that advances the state of multimodal perception in autonomous navigation and sets the foundation for further exploration into intelligent sensor-driven mobility systems.

**Author Contributions:** Arvind Kumar Bhardwaj conceptualized the research idea, designed the SensorFusionNet architecture, and supervised the overall study execution. Murtuza Ahamed Khan was responsible for integrating and preprocessing the multi-sensor data, and conducting model training and validation. Maloth Bhavsingh contributed to the real-time implementation, performance evaluation under dynamic conditions, and supported the interpretation of results. All authors collaboratively contributed to manuscript drafting, critical revisions, and approved the final version for publication.

**Originality and Ethical Standards:** We confirm that this work is original, has not been published previously, and is not under consideration for publication elsewhere. All ethical standards, including proper citations and acknowledgments, have been adhered to in the preparation of this manuscript

**Data availability:** Data available upon request.

**Conflict of Interest:** There is no conflict of Interest.

**Ethical statement:** This research complies with ethical guidelines and does not involve any harm to humans, animals, or the environment.

**Funding:** The research received no external funding.

**Similarity checked:** Yes.

## References

- [1] Atefi, Y. Ge, S. Pitla, and J. Schnable, "Robotic technologies for high-throughput plant phenotyping: Contemporary reviews and future perspectives," *Frontiers in Plant Science*, vol. 12, p. 611940, 2021.
- [2] H. S. Choi, J. Jeong, Y. H. Cho, K. J. Yoon, and J. H. Kim, "Semantics-guided transformer-based sensor fusion for improved waypoint prediction," *arXiv preprint arXiv:2308.02126*, 2023.
- [3] H. S. Choi, J. Jeong, Y. H. Cho, K. J. Yoon, and J. H. Kim, "Cognitive TransFuser: Semantics-Guided Transformer-Based Sensor Fusion for Improved Waypoint Prediction," in *Proc. Int. Conf. Robot Intelligence Technol. Appl.*, Cham: Springer Nature Switzerland, Dec. 2023, pp. 1–12.
- [4] R. Dainelli, P. Toscano, S. F. Di Gennaro, and A. Matese, "Recent advances in unmanned aerial vehicle forest remote sensing—A systematic review. Part I: A general framework," *Forests*, vol. 12, no. 3, p. 327, 2021.
- [5] M. S. Lakshmi, K. S. Ramana, G. Ramu, K. Shyam Sunder Reddy, C. Sasikala, and G. Ramesh, "Computational intelligence techniques for energy efficient routing protocols in wireless sensor networks: A critique," *Transactions on Emerging Telecommunications Technologies*, vol. 35, no. 1, Nov. 2023, doi: 10.1002/ett.4888.
- [6] I. Dronova, C. Kislik, Z. Dinh, and M. Kelly, "A review of unoccupied aerial vehicle use in wetland applications: Emerging opportunities in approach, technology, and data," *Drones*, vol. 5, no. 2, p. 45, 2021.
- [7] S. Ecke et al., "UAV-based forest health monitoring: A systematic review," *Remote Sensing*, vol. 14, no. 13, p. 3205, 2022.

- [8] E. J. S. Headley, "Signal Processing and Data Analysis for Real-Time Intermodal Freight Classification through a Multimodal Sensor System," M.S. thesis, Univ. of Arkansas, 2021.
- [9] S. Chappidi and A. Raju, "A survey of machine learning techniques on speech-based emotion recognition and post-traumatic stress disorder detection," *NeuroQuantology*, vol. 20, no. 14, pp. 69–79, Oct. 2022, doi: 10.4704/nq.2022.20.14.NQ88010.
- [10] H. Lee, "Deep Convolutional Neural Network Based Single Tree Detection Using Volumetric Module From Airborne Lidar Data," 2022.
- [11] X. Liu, "Feature Detection from Mobile LiDAR Using Deep Learning," M.S. thesis, Miami Univ., 2019.
- [12] N. H. H. Marcelis, "Pedestrian Detection and Tracking for Mobile Robots in Human Environments," [Thesis/Institution not specified].
- [13] T. McDonald, M. Robinson, and G. Y. Tian, "Developments in 3D visualisation of the rail tunnel subsurface for inspection and monitoring," *Applied Sciences*, vol. 12, no. 22, p. 11310, 2022.
- [14] A. Swetha, M. S. Lakshmi, and M. R. Kumar, "Chronic kidney disease diagnostic approaches using efficient artificial intelligence methods," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 10, no. 1s, pp. 254–259, 2022. [Online]. Available: <https://www.ijisae.org/index.php/IJISAE/article/view/2289>.
- [15] M. Mousa-Pasandi, T. Liu, Y. Massoud, and R. Laganière, "RGB-LiDAR fusion for accurate 2D and 3D object detection," *Machine Vision and Applications*, vol. 34, no. 5, p. 86, 2023.
- [16] G. Pradeep, S. Ramamoorthy, M. Krishnamurthy, and V. Saritha, "Energy prediction and task optimization for efficient IoT task offloading and management," *Int. J. Intell. Syst. Appl. Eng.*, vol. 12, no. 1s, pp. 411–427, 2023.
- [17] G. Pradeep, T. Sreedevi, and S. Ramamoorthy, "Enhancing IoT security: DL-based task offloading," *Int. J. Comput. Eng. Res. Trends (IJCERT)*, vol. 10, no. 4, 2022.
- [18] G. Pradeep, T. Sreedevi, and S. Ramamoorthy, "Deep learning approaches for ensuring secure task scheduling in IoT systems," *Int. J. Comput. Eng. Res. Trends (IJCERT)*, vol. 10, no. 4, 2021.
- [19] G. Pradeep, P. Satyanaryana, and S. Ramamoorthy, "A novel cryptographic protocol for secure data sharing in cloud computing," *Int. J. Comput. Eng. Res. Trends*, vol. 10, no. 11, pp. 10–18, Nov. 2023.
- [20] G. Rizzoli, F. Barbato, and P. Zanuttigh, "Multimodal semantic segmentation in autonomous driving: A review of current approaches and future perspectives," *Technologies*, vol. 10, no. 4, p. 90, 2022.
- [21] S. Chappidi and A. Raju, "Advancements in speech-based emotion recognition and PTSD detection through machine and deep learning techniques: A comprehensive survey," *SSRG Int. J. Electron. Commun. Eng.*, vol. 11, no. 5, 2023, doi: 10.14445/23488549/IJECE-V11I5P121.
- [22] G. Singh et al., "ROAD: The road event awareness dataset for autonomous driving," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 1, pp. 1036–1054, 2022.
- [23] Sun, "Operational design domain monitoring and augmentation for autonomous driving," 2022.
- [24] H. Xu et al., "A survey on UAV applications in smart city management: Challenges, advances, and opportunities," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 8982–9010, 2023.
- [25] J. Zhang, "Scene Understanding for Intelligent Transportation and Mobility Assistance Systems," Ph.D. dissertation, Karlsruhe Institut für Technologie (KIT), Germany, 2023.
- [26] K. Zhou et al., "An information fusion based incipient fault diagnosis method for railway vehicle door system," *IEEE Trans. Intell. Vehicles*, vol. 9, no. 1, pp. 1320–1332, 2023.
- [27] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, "The German Traffic Sign Recognition Benchmark: A multi-class classification competition," *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, San Jose, CA, USA, 2011, pp. 1453–1460.: <https://benchmark.ini.rub.de/>